

BIRCH, STEWART, KOLASCH & BIRCH, LLP

TERRELL C BIRCH  
RAYMOND C STEWART  
JOSEPH A KOLASCH  
JAMES M SLATTERY  
BERNARD L SWEENEY\*  
MICHAEL K MUTTER  
CHARLES GORENSTEIN  
GERALD M MURPHY, JR  
LEONARD R SVENSSON  
TERRY L CLARK  
ANDREW D MEIKLE  
MARC S WEINER  
JOE MCKINNEY MUNCY  
ROBERT J KENNEY  
DONALD J DALEY  
JOHN W BAILEY  
JOHN A CASTELLANO  
GARY D YACURA

OF COUNSEL  
HERBERT M BIRCH (1905-1996)  
ELLIOT A GOLDBERG\*  
WILLIAM L GATES\*  
EDWARD H VALANCE  
RUPERT J BRADY (RET)\*  
F. PRINCE BUTLER  
FRED S WHISENHUNT

\*ADMITTED TO A BAR OTHER THAN VA

INTELLECTUAL PROPERTY LAW  
8110 GATEHOUSE ROAD  
SUITE 500 EAST  
FALLS CHURCH, VA 22042-1210  
USA

(703) 205-8000

FAX (703) 205-8050  
(703) 698-8590 (G IV)

e-mail: mailroom@bskb.com  
web: http://www.bskb.com

CALIFORNIA OFFICE  
COSTA MESA, CALIFORNIA

THOMAS S AUCHTERLON  
JAMES T ELLER, JR  
SCOTT L LOWE  
MARK J NUEL, PH D  
D RICHARD ANDERSON  
PAUL C LEWIS  
MARK W MILSTEAD\*  
RICHARD J GALLAGHER  
JAYNE M SAYDAH\*  
MARYANNE ARMSTRONG, PH.D  
HYUNG N SOHN  
ALAN PEDERSEN-GILES  
KECIA J REYNOLDS

REG PATENT AGENTS  
FREDERICK R HANDREN  
MAKI HATSUMI  
MIKE S RYU  
CRAIG A McROBBIE  
GARTH M DAHLEN, PH D  
ROBERT E GOOZNER, PH D  
MATTHEW J LATTIG  
TIMOTHY R WYCKOFF  
KRISTI L RUPERT, PH D  
LARRY J HUME  
ALBERT K LEE  
HRAIR A SAYADIAN, PH D  
EVE L FRANK, PH D  
MATTHEW T SHANLEY



Date: November 7, 2000  
Docket No.: 1163-0301P

Assistant Commissioner for Patents  
Box PATENT APPLICATION  
Washington, D.C. 20231

Sir:

Transmitted herewith for filing is the patent application of

Inventor(s): TASAKI, Hirohisa  
YAMAURA, Tadashi

For: SPEECH CODING APPARATUS AND SPEECH DECODING APPARATUS

Enclosed are:

X A specification consisting of 80 pages

X 16 sheet(s) of Formal drawings

X An assignment of the invention

X Certified copy of Priority Document(s)

X Executed Declaration X Original      Photocopy

     A verified statement to establish small entity status under 37  
CFR 1.9 and 37 CFR 1.27

     Preliminary Amendment

X Information Disclosure Statement, PTO-1449 and reference(s)

Other \_\_\_\_\_

The filing fee has been calculated as shown below:

LARGE ENTITY				SMALL ENTITY	
FOR	NO. FILED	NO. EXTRA	RATE FEE		RATE FEE
BASIC FEE	***** ***** *****	***** ***** *****	***** ***** \$710.00 *****	or	***** ***** \$355.00 *****
TOTAL CLAIMS	15 - 20 =	0	x18 =\$ 0.00	or	x 9 = \$ 0.00
INDEPENDENT	6 - 3 =	3	x80 =\$ 240.00	or	x 40 = \$ 0.00
MULTIPLE DEPENDENT CLAIM PRESENTED <u>no</u>			+270 = \$ 0.00	or	+135 = \$ 0.00
TOTAL \$ 950.00				TOTAL \$ 0.00	

X A check in the amount of \$ 990.00 to cover the filing fee and recording fee (if applicable) is enclosed.

\_\_\_\_\_ Please charge Deposit Account No. 02-2448 in the amount of \$ \_\_\_\_\_. A triplicate copy of this transmittal form is enclosed.

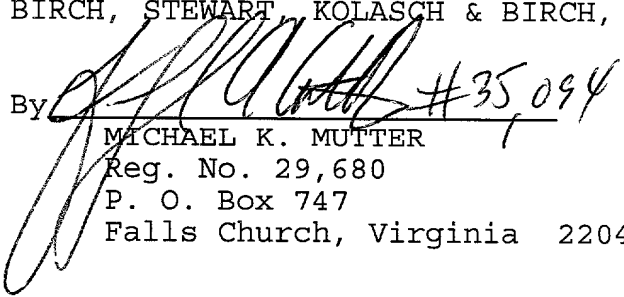
\_\_\_\_\_ No fee is enclosed.

If necessary, the Commissioner is hereby authorized in this, concurrent, and future replies, to charge payment or credit any overpayment to Deposit Account No. 02-2448 for any additional fees required under 37 C.F.R. 1.16 or under 37 C.F.R. 1.17; particularly, extension of time fees.

Respectfully submitted,

BIRCH, STEWART, KOLASCH & BIRCH, LLP

By

 #35,094  
MICHAEL K. MUTTER

Reg. No. 29,680

P. O. Box 747

Falls Church, Virginia 22040-0747

## SPEECH CODING APPARATUS AND SPEECH DECODING APPARATUS

## BACKGROUND OF THE INVENTION

## Field of the Invention

5       The present invention relates to a speech coding apparatus for compressing a digital speech signal to an equivalent signal having a smaller amount of information, and a speech decoding apparatus for decoding speech code generated by the speech coding apparatus or the like to reconstruct a  
10   digital speech signal.

## Description of the Prior Art

      Prior art speech coding apparatuses separate an input speech into spectral envelope information and an excitation source and encode them on a frame-by-frame basis, where each  
15   frame has a certain length, so as to generate speech code, and prior art speech decoding apparatuses decode the speech code and generate decoded speech by combining the spectral envelope information and the excitation source using a synthesis filter. Typical prior art speech coding apparatuses and speech decoding  
20   apparatuses employ a code-excited linear prediction (CELP) coding technique.

      Referring now to Fig. 14, there is illustrated a block diagram showing the structure of a prior art CELP speech coding apparatus. Fig. 15 is a block diagram showing the structure  
25   of a prior art CELP speech decoding apparatus. In Fig. 14, reference numeral 1 denotes an input speech, numeral 2 denotes a linear prediction analyzer, numeral 3 denotes a linear prediction coefficient coding unit, numeral 4 denotes an adaptive excitation source coding unit, numeral 5 denotes a  
30   driving excitation source coding unit, numeral 6 denotes a gain

coding unit, numeral 7 denotes a multiplexer, and numeral 8 denotes speech code. In Fig. 15, reference numeral 9 denotes a separator, numeral 10 denotes a linear prediction coefficient decoding unit, numeral 11 denotes an adaptive excitation source decoding unit, numeral 12 denotes a driving excitation source decoding unit, numeral 13 denotes a gain decoding unit, numeral 14 denotes a synthesis filter, and numeral 15 denotes output speech.

In operation, the prior art speech coding apparatus performs its coding operation on a frame-by-frame basis, where each frame has a duration ranging from 5 to 50 msec. Similarly, the prior art speech decoding apparatus performs its decoding operation on a frame-by-frame basis. In the speech coding apparatus of Fig. 14, the input speech 1 is applied to the linear prediction analyzer 2, the adaptive excitation source coding unit 4, and the gain coding unit 6. The linear prediction analyzer 2 analyzes the input speech 1 so as to extract a linear prediction coefficient that is the spectral envelope information of the input speech 1. The linear prediction coefficient coding unit 3 then encodes the linear prediction coefficient and furnishes the coded result to the multiplexer 7. The linear prediction coefficient coding unit 3 also quantizes the linear prediction and furnishes the quantized linear prediction to the adaptive excitation source coding unit 4, the driving excitation source coding unit 5, and the gain coding unit 6 for coding an excitation source separated from the input speech 1.

The adaptive excitation source coding unit 4 stores a past excitation source (or signal) of a certain length as an adaptive excitation source code book (i.e., adaptive code book) and

generates a plurality of adaptive excitation source codes each of which is a multiple-bit binary value. For each of the plurality of adaptive excitation source codes, the adaptive excitation source coding unit 4 also generates a time-series vector that is a series of pitch-cycles each of which includes the past excitation source. The adaptive excitation source coding unit 4 then multiplies the plurality of time-series vectors by an appropriate gain and allows the multiplication result to pass through a synthesis filter (not shown) using the quantized linear prediction coefficient from the linear prediction coefficient coding unit 3 so as to generate a temporary synthesized speech. The adaptive excitation source coding unit 4 calculates and examines the distance between the temporary synthesized speech and the input speech 1 and selects one adaptive excitation source code which minimizes the distance from the plurality of adaptive excitation source codes. The adaptive excitation source coding unit 4 then delivers the selected adaptive excitation source code to the multiplexer 7. The adaptive excitation source coding unit 4 also furnishes the time-series vector associated with the selected adaptive excitation source code as an adaptive excitation source to the driving excitation source coding unit 5 and the gain coding unit 6. The adaptive excitation source coding unit 4 further delivers either the input speech 1 or a signal obtained by substituting synthesized speech generated from the adaptive excitation source from the input signal 1, as a signal to be coded, to the driving excitation source coding unit 5.

The driving excitation source coding unit 5 contains a driving excitation source code book and generates a plurality of driving excitation source codes each of which is a

multiple-bit binary value. For each of the plurality of driving excitation source codes, the driving excitation source coding unit 5 also reads a time-series vector from the driving excitation source code book. The driving excitation source coding unit 5 then multiplies both the plurality of time-series vectors and the adaptive excitation source output from the adaptive excitation source coding unit 4 by respective appropriate gains and calculates the sum of them and allows the sum to pass through a synthesis filter (not shown) using the quantized linear prediction coefficient from the linear prediction coefficient coding unit 3 so as to generate a temporary synthesized speech. The driving excitation source coding unit 5 calculates and examines the distance between the temporary synthesized speech and the signal to be coded, which is either the input speech 1 or the signal obtained by substituting the synthesized speech generated from the adaptive excitation source from the input signal 1, and selects one driving excitation source code which minimizes the distance from the plurality of driving excitation source codes. The driving excitation source coding unit 5 then delivers the selected driving excitation source code to the multiplexer 7. The driving excitation source coding unit 5 also furnishes the time-series vector associated with the selected driving excitation source code as a driving excitation source to the gain coding unit 6.

The gain coding unit 6 stores a gain code book therein and generates a plurality of gain codes, each of which is a multiple-bit binary value. For each of the plurality of gain codes, the gain coding unit 6 also reads a gain vector sequentially from the gain code book. The gain coding unit 6

then multiplies both the adaptive excitation source output from the adaptive excitation source coding unit 4 and the driving excitation source output from the driving excitation source coding unit 5 by two elements of the gain vector, respectively, and calculates the sum of them so as to generate an excitation source and allows the excitation source to pass through a synthesis filter (not shown) using the quantized linear prediction coefficient from the linear prediction coefficient coding unit 3 so as to generate a temporary synthesized speech.

10 The gain coding unit 6 calculates and examines the distance between the temporary synthesized speech and the input speech 1, and selects one gain code which minimizes the distance from the plurality of gain codes. The gain coding unit 6 then delivers the selected gain code to the multiplexer 7. The gain coding unit 6 also furnishes the generated excitation source corresponding to the selected gain code to the adaptive excitation source coding unit 4.

Finally, the adaptive excitation source coding unit 4 updates the adaptive code book located therein using the excitation source corresponding to the gain code selected by the gain coding unit 6.

The multiplexer 7 multiplexes the linear prediction coefficient code from the linear prediction coefficient coding unit 3, the adaptive excitation source code from the adaptive excitation source coding unit 4, the driving excitation source code from the driving excitation source coding unit 5, and the gain code from the gain coding unit 6 into a speech code 8, and outputs the speech code 8.

In the speech decoding apparatus of Fig. 15, the separator 9 separates the speech code 8 from the speech coding apparatus

into the linear prediction coefficient code, the adaptive  
excitation source code, the driving excitation source code, and  
the gain code. The separator 9 then furnishes them to the linear  
prediction coefficient decoding unit 10, the adaptive  
5 excitation source decoding unit 11, the driving excitation  
source decoding unit 12, and the gain decoding unit 13,  
respectively. The linear prediction coefficient decoding unit  
10 decodes the linear prediction coefficient code from the  
separator 9 so as to reconstruct the linear prediction  
10 coefficient. The linear prediction coefficient decoding unit  
10 then sets and outputs the linear prediction coefficient as  
a filter coefficient for the synthesis filter 14.

The adaptive excitation source decoding unit 11 stores  
a past excitation source as an adaptive excitation source code  
15 book. The adaptive excitation source decoding unit 11 also  
generates a time-series vector that is a series of pitch-cycles  
each of which includes the past excitation source, as an  
adaptive excitation source, the time-series vector being  
associated with the adaptive excitation source code separated  
20 by the separator 9. The driving excitation source decoding unit  
12 generates a time-series vector as a driving excitation source,  
the time-series vector being associated with the driving  
excitation source code separated by the separator 9. The gain  
decoding unit 13 also generates a gain vector associated with  
25 the gain code separated by the separator 9. The speech decoding  
apparatus then multiplies both the first and second time-series  
vectors from the adaptive excitation source decoding unit and  
the driving excitation source decoding unit by two elements of  
the gain vector from the gain decoding unit, respectively, so  
30 as to generate an excitation source and allows the excitation



source to pass through the synthesis filter 14 so as to generate output speech 15. Finally, the adaptive excitation source decoding unit 11 updates the adaptive excitation source code book located therein using the generated excitation source.

5       Next, a description will be made as to an improvement in the prior art CELP speech coding and decoding apparatuses mentioned above. "Basic algorithm of conjugate-structure algebraic CELP (CS-ACELP) speech coder" by A. Kataoka et al., NTT R&D, Vol. 45, April 1996, which will be referred to as  
10   Reference 1, discloses a CELP speech coding apparatus and a CELP speed decoding apparatus including a excitation source pulse for coding a driving excitation source with the aim of reducing the amount of calculations and the amount of memory. In this prior art arrangement, the driving excitation source is  
15   represented only by information about the locations of a number of pulses and information about the polarities of the plurality of pulses. Such an excitation source is called an algebraic excitation source, and provides a good coding performance considering that it has a simple structure. Recently-  
20   developed standard coding techniques adopt the algebraic excitation source.

Referring next to Fig. 16, there is illustrated a table listing candidates for the locations of the excitation source pulses employed by the CELP speech coding and decoding  
25   apparatuses disclosed in Reference 1. Such the table can be located in both the driving excitation source coding unit 5 of the speech coding apparatus as shown in Fig. 14 and the driving excitation source decoding unit 12 of the speech decoding apparatus as shown in Fig. 15. In Reference 1, the length of  
30   frames to be coded when coding excitation sources is 40 samples,

and the driving excitation source consists of four pulses. Three of them numbered 1 to 3 have 8 limited possible locations as shown in Fig. 16, respectively. Therefore, each of the locations of the three pulses can be coded in three bits. The  
 5 remaining pulse numbered 4 has 16 limited possible locations as shown in Fig. 16. Therefore, the location of the fourth pulse can be coded in four bits. The number of candidates for the location of each of the four excitation source pulses is limited in this way, and the amount of bits used for coding the driving  
 10 excitation source and the number of combinations of the locations of those excitation source pulses are therefore reduced. This results in a reduction in the amount of arithmetic operations without reducing the coding performance.

In accordance with the coding technique as disclosed in  
 15 Reference, the driving excitation source coding unit 5 of the speech coding apparatus of Fig. 14 calculates a correlation between an impulse response (i.e., a synthesized speech generated by a single excitation source pulse) and a signal to be coded, and a cross-correlation between impulse responses  
 20 (i.e., synthesized speeches respectively generated by single excitation source pulses), and stores them as a pre-table therein and calculates the distance (or coding distortion) by simply calculating the sum of them. The driving excitation source coding unit 5 then searches for the pulse locations and  
 25 polarities that minimize the distance.

The concrete searching method as disclosed in Reference 1 will be described hereinafter. The minimization of the distance is equivalent to the maximization of an evaluation value  $D$  given by the following equation:

$$D = C^2 / E \quad (1)$$

where C and E are given by:

$$C = \sum_k g(k) d(m_k) \quad (2)$$

$$E = \sum_k \sum_i g(k) g(i) \phi(m_k, m_i) \quad (3)$$

5

where  $m_k$  is the location of the  $k$ th pulse,  $g(k)$  is the magnitude of the  $k$ th pulse,  $d(x)$  is the correlation between an impulse response generated when an impulse is placed at the pulse position  $x$  and the signal to be coded, and  $\phi(x, y)$  is the cross-correlation between an impulse response generated when an impulse is placed at the pulse location  $x$  and an impulse response generated when an impulse is placed at the pulse location  $y$ . The searching process is carried out by the calculation of the evaluation value  $D$  for all combinations of the possible locations of all excitation source pulses.

In addition, simplifying the above equations (2) and (3) by assuming that  $g(k)$  has the same sign as  $d(m_k)$  and has an absolute value of 1 yields the following equations (4) and (5):

$$C = \sum_k d'(m_k) \quad (4)$$

$$E = \sum_k \sum_i \phi'(m_k, m_i) \quad (5)$$

20

where

$$d'(m_k) = |d(m_k)| \quad (6)$$

$$\phi'(m_k, m_i) = \text{sign}[d(m_k)] \text{sign}[d(m_i)] \phi(m_k, m_i) \quad (7)$$

Only calculating  $d'(m_k)$  and  $\phi'(m_k, m_i)$  in advance of the calculation of the evaluation value D for all combinations of the locations of all excitation source pulses is thus needed before the simple summations according to the equations (4) and (5), thereby reducing the amount of arithmetic operations.

Japanese patent application publications (TOKKAIHEI) No. 10-232696 and No.10-312198, and "Improvements in ACELP speech coding based on adaptive pulse locations", by Tsuchiya et al., Nihon Onkyo Gakkai (The Acoustical Society of Japan) 1999 Shunki Kenkyuu Happyokai Kouen Ronbunshuu vol.I, pp.213-214, 1999, which will be referred to as Reference 2, disclose configurations for improving the quality of the algebraic excitation source mentioned above.

Japanese patent application publication No. 10-232696 discloses a method of providing a plurality of fixed waveforms and generating a driving excitation source by placing the plurality of fixed waveforms at a plurality of locations coded algebraically, respectively, thereby yielding an output speech with a high quality. Reference 2 studies an arrangement in which a pitch filter is contained in a generating unit for generating a driving excitation source (in reference 2, an ACELP excitation source). Either of the arrangement of the plurality of fixed waveforms and the pitch-filtering process to generate a pitch-filtered driving excitation source can improve the quality of the output speech without increasing the amount of searching operations if it is carried out at the same time that the calculation of impulse responses is done.

Japanese patent application publication No. 10-312198

discloses an arrangement in which the locations of excitation  
sources pulses are searched for while the driving excitation  
source is made to be orthogonal to the adaptive excitation  
source when the pitch gain is greater than or equal to a  
5 predetermined value.

Referring next to Fig. 17, there is illustrated a block  
diagram showing in details the structure of a driving excitation  
source coding unit 5 of an improved CELP speech coding apparatus  
disclosed in Japanese patent application publication No.  
10 10-232696 and Reference 2. In the figure, reference numeral  
16 denotes a perceptual weighting filter coefficient  
calculating unit, numerals 17 and 19 denote perceptual  
weighting filters, numeral 18 denotes a basic response  
generating unit, numeral 20 denotes a pre-table calculating  
15 unit, numeral 21 denotes a searching unit, and numeral 22  
denotes an excitation source location table.

Next, the operation of the driving excitation source  
coding unit 5 will be described. A quantized linear prediction  
coefficient from a linear prediction coefficient coding unit  
20 3 disposed within the speech coding apparatus as shown in Fig.  
14 is applied to the perceptual weighting filter coefficient  
calculating unit 16 and the basic response generating unit 18.  
An adaptive excitation source coding unit 4 furnishes a signal  
to be coded that is either an input speech 1 or a signal obtained  
25 by substituting synthesized speech generated from an adaptive  
excitation source from the input speech 1 to the perceptual  
weighting filter 17. The adaptive excitation source coding  
unit 4 also delivers the repetition period of the adaptive  
excitation source converted from an adaptive excitation source  
30 code to the basic response generating unit 18.

The perceptual weighting filter coefficient calculating unit 16 then calculates a perceptual weighting filter coefficient using the quantized linear prediction coefficient and sets the calculated perceptual weighting filter coefficient as a filter coefficient intended for the perceptual weighting filters 17 and 19. The perceptual weighting filter 17 performs a filtering process on the input signal to be coded using the filter coefficient set by the perceptual weighting filter coefficient calculating unit 16.

The basic response generating unit 18 performs pitch filtering on a unit impulse or a fixed waveform using the repetition period of the adaptive excitation source so as to generate a series of cycles each of which includes the unit impulse or the fixed waveform, the repetition period of the series of cycles being equal to that of the adaptive excitation source. The basic response generating unit 18 then allows the generated signal, as an excitation source, to pass through a synthesis filter formed using the quantized linear prediction coefficient to generate synthesized speech, and outputs the synthesized speech as a basic response. The perceptual weighting filter 19 performs a filtering process on the basis response using the filter coefficient set by the perceptual weighting filter coefficient calculating unit 16.

The pre-table calculating unit 20 calculates the correlation  $d(x)$  between the perceptual weighted signal to be coded and the perceptual weighted basic response when placing the impulse at the location  $x$ , and calculates the cross-correlation  $\phi(x,y)$  between the perceptual weighted basic response when placing the impulse at the location  $x$  and the perceptual weighted basic response when placing the impulse at

the location  $y$ . The pre-table calculating unit 20 then obtains  $d'(x)$  and  $\phi'(x,y)$  according to equations (6) and (7) and stores them as a pre-table.

The excitation source location table 22 stores a plurality of candidates for the locations of excitation source pulses, which are similar to those as shown in Fig. 16. The searching unit 21 sequentially reads each of all combinations of the possible locations of the excitation source pulses from the excitation source location table 22 and calculates an evaluation value  $D$  for each combination of the possible locations of the excitation source pulses using the pre-table calculated by the pre-table calculating unit 20 according to above-mentioned equations (1), (4) and (5). The searching unit 21 also searches for one combination of the possible locations of the excitation source pulses which maximizes the evaluation value  $D$  and furnishes excitation source location code (i.e., indexes of the excitation source location table) indicating the combination of the possible locations of the excitation source pulses and polarity code indicating the polarities of them, as driving excitation source code, to a multiplexer 7 as shown in Fig. 14. The searching unit 21 further delivers one time-series vector associated with the driving excitation source code to a gain coding unit 6 as shown in Fig. 14.

In Japanese patent application publication No. 10-312198, the method of making the driving excitation source orthogonal to the adaptive excitation source is implemented by making the perceptual weighted signal to be coded which is input to the pre-table calculating unit 20 orthogonal to the adaptive excitation source, and contributions associated with the correlation between the adaptive excitation source and each

driving excitation source pulse are subtracted from E given by equation (5) in the searching unit 21.

A problem encountered with prior art speech coding apparatuses and prior art speech decoding apparatuses constructed as above is that while the pitch-filtering process to generate a pitch-filtered driving excitation source can improve the coding performance without increasing the amount of searching operations, the use of the repetition period of an adaptive excitation source as the repetition period intended for the pitch-filtering process can degrade the quality of speech code generated when the pitch-period of an input speech is different from the repetition period of the adaptive excitation source.

Fig. 18 shows a relationship between a signal to be coded and the locations of pulses included in each pitch-cycle of a pitch-filtered driving excitation source, when the repetition period of the adaptive excitation source is two times the pitch-period of an input speech, in accordance with a prior art speech coding apparatus and a prior art speech decoding apparatus. Fig. 19 shows a relationship between a signal to be coded and the locations of pulses included in each pitch-cycle of a pitch-filtered driving excitation source, when the repetition period of the adaptive excitation source is one-half the pitch-period of an input speech, in accordance with a prior art speech coding apparatus and a prior art speech decoding apparatus.

The repetition period of the adaptive excitation source is determined such that the coding distortion between a synthesized speech generated based on the adaptive excitation source and the signal to be coded is minimized. Therefore the



repetition period of the adaptive excitation source is frequently different from the pitch-period of the input speech that is the period of vibrations of the speaker's vocal cords. In this case, the repetition period of the adaptive excitation source is approximately an integral multiple or submultiple of the pitch-period of the input speech. In many cases, the repetition period of the adaptive excitation source is about two times or one-half the pitch-period.

In Fig. 18, since the speaker's vocal cords vibrate in the same way every other pitch-cycle, it is determined that the repetition period of the adaptive excitation source is about two times as large as the pitch-period of the input speech. When the driving excitation source is coded using the repetition period of the adaptive excitation source, most excitation source pulses are concentrated in the first half of the period of each pitch-cycle. The pitch-filtered driving excitation source that is the series of pitch-cycles thus obtained in the current frame using the repetition period of the adaptive excitation source is as shown in Fig. 18. The use of the excitation source pitch-filtered using the repetition period different from the pitch-period of the input speech can cause a change in the tone quality of the frame and hence instability in the synthesized speech. This disadvantage does not become negligible as the bit rate decreases and the amount of information about the driving excitation source therefore decreases. Frames in which the magnitude of the adaptive excitation source is less than that of the driving excitation source have noticeable degradation of the sound quality.

In Fig. 19, since there is a predominance of low-frequency components in the input speech signal and the waveform of the

first half of each pitch-cycle of the input speech is similar to that of the second half of each pitch-cycle, it is determined that the repetition period of the adaptive excitation source is about one-half the pitch-period of the input speech. As in  
5 the case of Fig. 18, the use of the excitation source pitch-filtered using the repetition period different from the pitch-period of the input speech can cause a change in the tone quality of the frame and hence unstability in the synthesized speech.

10 When the bit rate decreases and the amount of information about the driving excitation source therefore decreases, there is a tendency that the driving excitation source determined such that the waveform distortion (or coding distortion) is minimized has a large error in a band of low magnitudes and the  
15 synthesized speech therefore has a large spectral distortion. Such a spectral distortion can be detected as degradation of the sound quality. Although a perceptual weighting process is provided in order to eliminate degradation of the sound quality due to spectral distortions, an enhancement of the perceptual  
20 weighting process can cause an increase in the waveform distortion and hence degradation of the sound quality showing a ragged sound. The enhancement of the perceptual weighting process is therefore controlled such that the adverse effect on the sound quality by the waveform distortion has the same  
25 level as that by the spectral distortion. However, the spectral distortion is increased when the input speech is a female one, and the perceptual weighting process cannot be controlled so that it is optimized for both male and female speeches.

In prior art configurations, a constant magnitude is  
30 provided for a plurality of excitation sources, such as pulses,

placed at respective locations within each pitch-cycle included in each frame. There is no use in equalizing the magnitudes of the plurality of excitation sources regardless of the difference in the number of candidates for the location of each of the plurality of excitation sources. In the excitation source location table as shown in Fig. 16, three bits are used for each of the excitation source locations numbered 1 to 3 and four bits are used for the remaining excitation source location numbered 4. It is easily expected by examining a maximum of a correlation between each of the plurality of excitation sources placed at a possible location and the signal to be coded that the excitation source number 4 having the largest number of possible locations has a higher probability of providing the largest correlation. Assume an extreme case where no bit is provided for an excitation source number. In the case where no bit is provided for an excitation source number, i.e., one excitation source is fixed at a certain location, the correlation between the excitation source and the signal to be coded is small while the polarity is provided independently. This means that it is not appropriate to provide a larger magnitude for one excitation source as compared with those provided for other excitation sources. The problem with prior art configurations is thus that the magnitudes of the plurality of excitation sources are not optimized.

Although a prior art configuration is disclosed for providing an individual magnitude for each of the plurality of excitation sources through vector quantization during the gain quantization process, the amount of gain-quantized information increases and the gain quantization process increases in complexity.

The above-mentioned technique of making the driving excitation source orthogonal to the adaptive excitation source causes an increase in the amount of searching operations. Therefore, an increase in the number of combinations of algebraic excitation sources puts an enormous load on the coding or decoding process. Especially, when using the technique of making the driving excitation source orthogonal to the adaptive excitation source in a prior art configuration that generates a driving excitation source by placing a plurality of fixed waveforms or performs a pitch-filtering process to generate a pitch-filtered driving excitation source, the amount of arithmetic operations increase greatly.

#### SUMMARY OF THE INVENTION

The present invention is proposed to solve the above problems. It is therefore an object of the present invention to provide a speech coding apparatus capable of generating high-quality speech code and a speech decoding apparatus capable of reconstructing a high-quality speech.

It is another object of the present invention to provide a speech coding apparatus capable of generating high-quality speech code while keeping an increase in the amount of arithmetic operations to a minimum and a speech decoding apparatus capable of reconstructing a high-quality speech while keeping an increase in the amount of arithmetic operations to a minimum.

In accordance with one aspect of the present invention, there is provided a speech coding apparatus for coding an input speech on a frame-by-frame basis using an adaptive excitation source, which is generated from a past excitation source, and

a driving excitation source, which is generated from the input speech and the adaptive excitation source, so as to generate speech code, the speech coding apparatus comprising: a repetition period pre-selecting unit for generating a plurality of candidates for a repetition period of the driving excitation source by multiplying a repetition period of the adaptive excitation source by a plurality of constant numbers, respectively, and for pre-selecting a predetermined number of candidates from all the candidates generated and furnishing the predetermined number of pre-selected candidates; a driving excitation source coding unit for providing both excitation source location information and excitation source polarity information that minimize a coding distortion, for each of the predetermined number of candidates for the repetition period of the driving excitation source, and for providing an evaluation value associated with the minimum coding distortion for each of the predetermined number of candidates; and a repetition period coding unit for comparing the evaluation values provided for the predetermined number of candidates for the repetition period of the driving excitation source from the driving excitation source coding unit with one another, for selecting one candidate from the predetermined number of candidates according to a comparison result, and for furnishing selection information indicating a selection result, excitation source location code indicating excitation source location information associated with the selected candidate for the repetition period of the driving excitation source, and polarity code indicating excitation source polarity information associated with the selected candidate.

In accordance with a preferred embodiment of the present

invention, the repetition period pre-selecting unit pre-selects two candidates from all the candidates generated, and the repetition period coding unit encodes the selection result in one bit so as to generate 1-bit selection information.

5 In accordance with another preferred embodiment of the present invention, the repetition period pre-selecting unit includes a unit for comparing the repetition period of the adaptive excitation source with a predetermined threshold value, and for pre-selecting the predetermined number of candidates  
10 from all the candidates generated according to a comparison result.

In accordance with another preferred embodiment of the present invention, the repetition period pre-selecting unit includes a unit for generating a plurality of other adaptive  
15 excitation sources whose respective repetition periods equal to the plurality of candidates for the repetition period of the driving excitation source, respectively, and for pre-selecting the predetermined number of candidates from all the candidates generated according to a comparison between distances among the  
20 plurality of other adaptive excitation sources generated.

Preferably, the plurality of constant numbers, by which the repetition period of the adaptive excitation source is multiplied, includes  $1/2$  and  $1$ .

In accordance with another aspect of the present  
25 invention, there is provided a speech decoding apparatus for decoding input speech code on a frame-by-frame basis using an adaptive excitation source, which is generated from a past excitation source, and a driving excitation source, which is generated from the input speech code and the adaptive excitation  
30 source, so as to reconstruct original speech, the speech

decoding apparatus comprising: a repetition period pre-selecting unit for providing a plurality of candidates for a repetition period of the driving excitation source by multiplying a repetition period of the adaptive excitation source by a plurality of constant numbers, respectively, and for pre-selecting a predetermined number of candidates from all the candidates generated and furnishing the predetermined number of pre-selected candidates; a repetition period decoding unit for selecting one candidate from the predetermined number of pre-selected candidates for the repetition period of the driving excitation source from the repetition period pre-selecting unit according to selection information included in the input coded speech and indicating the selection, and for furnishing the selected candidate as the repetition period of the driving excitation source; and a driving excitation source decoding unit for generating a time-series signal according to excitation source location code and excitation source polarity code included in the input speech code, and for generating a time-series vector that is a series of pitch-cycles, each of which includes the time-series signal, using the repetition period of the driving excitation source from the repetition period decoding unit.

In accordance with a preferred embodiment of the present invention, the repetition period pre-selecting unit pre-selects two candidates from all the candidates generated, and the repetition period decoding unit decodes selection information coded in one bit, which is included in the input speech code and indicates a selection of a candidate for the repetition period of the adaptive excitation source made during coding.

In accordance with another preferred embodiment of the present invention, the repetition period pre-selecting unit includes a unit for comparing the repetition period of the adaptive excitation source with a predetermined threshold value, and for pre-selecting the predetermined number of candidates from all the candidates generated according to a comparison result.

In accordance with another preferred embodiment of the present invention, the repetition period pre-selecting unit includes a unit for generating a plurality of other adaptive excitation sources whose respective repetition periods equal to the plurality of candidates for the repetition period of the driving excitation source, respectively, and for pre-selecting the predetermined number of candidates from all the candidates generated according to a comparison between distances among the plurality of other adaptive excitation sources generated.

Preferably, the plurality of constant numbers, by which the repetition period of the adaptive excitation source is multiplied, includes  $1/2$  and  $1$ .

In accordance with a further aspect of the present invention, there is provided a speech coding apparatus for coding an input speech on a frame-by-frame basis using an adaptive excitation source, which is generated from a past excitation source, and a driving excitation source, which is generated from the input speech and the adaptive excitation source, so as to generate speech code, the speech coding apparatus comprising: a perceptual weighting control unit for determining a perceptual weighting strength coefficient based on a repetition period of the adaptive excitation source; and a driving excitation source coding unit for generating



excitation source location code indicating information about excitation source locations and information about excitation source polarities based on the repetition period of the adaptive excitation source, the perceptual weighting strength

5 coefficient determined by the perceptual weighting control unit, and a signal to be coded such as the input speech.

In accordance with a preferred embodiment of the present invention, the perceptual weighting control unit determines the perceptual weighting strength coefficient based on an average  
10 of the repetition period of the current adaptive excitation source and repetition periods of previously-generated adaptive excitation sources.

In accordance with another aspect of the present invention, there is provided a speech coding apparatus for  
15 coding an input speech on a frame-by-frame basis using an adaptive excitation source, which is generated from a past excitation source, and a driving excitation source generated from the input speech and the adaptive excitation source, the driving excitation source being represented by locations and  
20 polarities of a plurality of excitation sources, so as to generate speech code, the speech coding apparatus comprising: an excitation source location table including a plurality of selectable possible locations and a fixed magnitude determined based on the number of the plurality of possible locations for  
25 each of the plurality of excitation sources; a driving excitation source coding unit for placing the plurality of excitation sources at respective possible locations while multiplying each of the plurality of excitation sources by a corresponding fixed magnitude, with reference to the excitation  
30 source location table, for generating a driving excitation

source by calculating a sum of the plurality of excitation sources each of which has been multiplied by the corresponding fixed magnitude and is thus placed at one corresponding possible location, for each of all combinations of possible locations of the plurality of excitation sources, and for selecting possible locations and polarities of the plurality of excitation sources which provide a driving excitation source having a smallest coding distortion between itself and the input speech so as to generate excitation source location code and polarity code.

In accordance with a further aspect of the present invention, there is provided a speech decoding apparatus for decoding input speech code on a frame-by-frame basis using an adaptive excitation source, which is generated from a past excitation source, and a driving excitation source generated from the input speech code and the adaptive excitation source, the driving excitation source being represented by locations and polarities of a plurality of excitation sources, so as to reconstruct original speech, the speech decoding apparatus comprising: an excitation source location table including a plurality of selectable possible locations and a fixed magnitude determined based on the number of the plurality of possible locations for each of the plurality of excitation sources; a driving excitation source decoding unit for selecting respective possible locations for the plurality of excitation sources with reference to the excitation source location table based on excitation source location code included in the input speech code, for placing the plurality of excitation sources at the respective selected possible locations while multiplying each of the plurality of excitation

sources by a corresponding fixed magnitude, and for generating a driving excitation source by calculating a sum of the plurality of excitation sources each of which has been multiplied by the corresponding fixed magnitude and is thus placed at the corresponding possible location.

In accordance with another aspect of the present invention, there is provided a speech coding apparatus for coding an input speech on a frame-by-frame basis using an adaptive excitation source, which is generated from a past excitation source, and a driving excitation source generated from the input speech and the adaptive excitation source, the driving excitation source being represented by locations and polarities of a plurality of excitation sources, so as to generate speech code, the speech coding apparatus comprising: a pre-table calculating unit for calculating a correlation between a signal to be coded, such as the input speech, and each of a plurality of synthesized speeches each of which is generated based on a corresponding temporary driving excitation source that is a signal obtained by placing a predetermined excitation source at a corresponding one of all possible locations, and a cross-correlation between any two of the plurality of synthesized speeches, and for storing these calculated correlations and cross-correlations as a pre-table therein; a pre-table modifying unit for calculating a correlation between the signal to be coded and a synthesized speech generated based on the adaptive excitation source, and a correlation between each of the plurality of synthesized speeches generated based on the corresponding temporary driving excitation source and the synthesized speech generated based on the adaptive excitation source, and for modifying the

pre-table using these calculated correlations; and a searching unit for determining the locations and polarities of the plurality of excitation sources using the pre-table corrected by the pre-table modifying unit so as to generate excitation source location code indicating the locations of the plurality of excitation sources and excitation source polarity code indicating the polarities of the plurality of excitation sources.

Further objects and advantages of the present invention will be apparent from the following description of the preferred embodiments of the invention as illustrated in the accompanying drawings.

#### BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a block diagram showing the structure of a driving excitation source coding unit of a speech coding apparatus according to a first embodiment of the present invention;

Fig. 2 is a block diagram showing the structure of a driving excitation source decoding unit of a speech decoding apparatus according to the first embodiment of the present invention;

Fig. 3 is a diagram showing a relationship between a signal to be coded and the locations of pulses of each of a series of cycles included in a cyclic adaptive excitation source, when the repetition period of the adaptive excitation source is two times the pitch-period of an input speech, in accordance with the first embodiment of the present invention;

Fig. 4 is a diagram showing a relationship between the signal to be coded and the locations of pulses of each of a series

of cycles included in a cyclic adaptive excitation source, when the repetition period of the adaptive excitation source is one-half the pitch-period of an input speech, in accordance with the first embodiment of the present invention;

5        Fig. 5 is a block diagram of a driving excitation source coding unit of a speech coding apparatus according to a second embodiment of the present invention;

10       Fig. 6 is a block diagram showing the structure of a driving excitation source decoding unit of a speech decoding apparatus according to the second embodiment of the present invention;

15       Fig. 7 is a diagram showing other adaptive excitation sources generated by an adaptive excitation source generating unit of the speech decoding apparatus according to the second embodiment of the present invention when the repetition period of an original adaptive excitation source is equal to the pitch-period of an input speech;

20       Fig. 8 is a diagram showing other adaptive excitation sources generated by the adaptive excitation source generating unit of the speech decoding apparatus according to the second embodiment of the present invention when the repetition period of an original adaptive excitation source is twice the pitch-period of an input speech;

25       Fig. 9 is a diagram showing other adaptive excitation sources generated by the adaptive excitation source generating unit of the speech decoding apparatus according to the second embodiment of the present when the repetition period of an original adaptive excitation source is three times the pitch-period of an input speech;

30       Fig. 10 is a block diagram showing the structure of a

driving excitation source coding unit and a perceptual weighting control unit disposed within a speech coding apparatus according to a third embodiment of the present invention;

5        Fig. 11 is a block diagram showing the structure of a driving excitation source coding unit and a perceptual weighting control unit disposed within a speech coding apparatus according to a fourth embodiment of the present invention;

10       Fig. 12 is a diagram showing an excitation source location table according to a fifth embodiment of the present invention;

      Fig. 13 is a block diagram showing the structure of a driving excitation source coding unit of a speech coding apparatus in accordance with a sixth embodiment of the present invention;

15

      Fig. 14 is a block diagram showing the structure of a prior art CELP speech coding apparatus;

      Fig. 15 is a block diagram showing the structure of a prior art CELP speech decoding apparatus;

20       Fig. 16 is a diagram showing candidates for the locations of prior art excitation source pulses;

      Fig. 17 is a block diagram showing in details the structure of a driving excitation source coding unit of a prior art CELP speech coding apparatus;

25       Fig. 18 is a diagram showing a relationship between a signal to be coded and the locations of pulses included in each pitch-cycle of a pitch-filtered driving excitation source, when the repetition period of the adaptive excitation source is two times the pitch-period of an input speech, in accordance with

30       a prior art speech coding apparatus and a prior art speech

decoding apparatus; and

Fig. 18 is a diagram showing a relationship between a signal to be coded and the locations of pulses included in each pitch-cycle of a pitch-filtered driving excitation source, when the repetition period of the adaptive excitation source is one-half the pitch-period of an input speech, in accordance with a prior art speech coding apparatus and a prior art speech decoding apparatus.

## 10 DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

### Embodiment 1

Referring next to Fig. 1, there is illustrated a block diagram showing the structure of a driving excitation source coding unit of a speech coding apparatus in accordance with a first embodiment of the present invention. The speech coding apparatus has the same overall structure as shown in Fig. 14. In Fig. 1, reference numeral 23 denotes a repetition period pre-selecting unit, numeral 27 denotes a driving excitation source coder, and numeral 28 denotes a repetition period coder. The repetition period pre-selecting unit 23 includes a constant number table 24, a comparator 25, and a pre-selecting unit 26.

The driving excitation source coding unit 5 of the speech coding apparatus of this embodiment thus includes the driving excitation source coder 27 that operates in the same way that the prior art driving excitation source coding unit as mentioned above does, and the repetition period pre-selecting unit 23 and the repetition period coder 28 disposed in the front and back of the driving excitation source coder 27.

Referring next to Fig. 2, there is illustrated a block diagram showing the structure of a driving excitation source

decoding unit of a speech decoding apparatus in accordance with the first embodiment of the present invention. The speech decoding apparatus has the same overall structure as shown in Fig. 15. In Fig. 2, reference numeral 29 denotes a repetition period decoder, and numeral 30 denotes a driving excitation source decoder.

The driving excitation source decoding unit 12 of the speech decoding apparatus of this embodiment thus includes the driving excitation source decoder 30 that operates in the same way that the prior art driving excitation source decoding unit as mentioned above does, and the repetition period pre-selecting unit 23 and the repetition period decoder 29 inserted in the front of the driving excitation source decoder 30.

Next, a description will be made as to the operation of the speech coding apparatus with reference to Fig. 1. An adaptive excitation source coding unit 4 can convert an adaptive excitation source code into the repetition period of an adaptive excitation source. The repetition period of the adaptive excitation source is then delivered to the repetition period pre-selecting unit 23. Both a signal to be coded from the adaptive excitation source coding unit 4 and a quantized linear prediction coefficient from a linear prediction coefficient coding unit 3 are input to the driving excitation source coder 27.

The constant number table 24 disposed within the repetition period pre-selecting unit 23 stores three constant numbers:  $1/2$ , 1, and 2. The input repetition period of the adaptive excitation source is multiplied by the three constant numbers, respectively, and the three multiplication results are furnished as three candidates for the repetition period of the



driving excitation source to the pre-selecting unit 26. The comparator 25 compares the three possible repetition periods of the driving excitation source with a predetermined threshold value, respectively, and furnishes the comparison results to the pre-selecting unit 26. An averaged pitch-period of about 40 can be used as the threshold value.

The pre-selecting unit 26 pre-selects the two possible repetition periods of the driving excitation source obtained by multiplying the input repetition period of the adaptive excitation source by  $1/2$  and  $1$  when the comparison results indicate that all the multiplication results are greater than the predetermined threshold value, and, otherwise, pre-selects the two possible repetition periods of the driving excitation source obtained by multiplying the input repetition period of the adaptive excitation source by  $1$  and  $2$ . The pre-selecting unit 26 then delivers the two selected possible repetition periods of the driving excitation source to the driving excitation source coder 27 sequentially.

Like the prior art driving excitation source coding unit as shown in Fig. 17, the driving excitation source coder 27 can encode the algebraic excitation source using the two possible repetition periods of the driving excitation source, the quantized linear prediction coefficient, and the signal to be coded, and provide the locations of a plurality of excitation sources that minimize the coding distortion, each of the plurality of excitation sources consisting of either a fixed waveform or a pulse, the polarities of the plurality of excitation sources, and an evaluation value  $D$  associated with the coding distortion according to equation (1) described above, for each of the two possible repetition periods of the driving

excitation source. The driving excitation source coder 27 differs from the prior art driving excitation source coding unit as shown in Fig. 17 in that each of the received candidates for the repetition period of the driving excitation source is the one obtained by multiplying the repetition period of the adaptive excitation source by a constant number.

The repetition period coder 28 compares the two evaluation values D obtained for the two possible repetition periods of the driving excitation source from the driving excitation source coder 27 with each other. If the difference between them is equal to or greater than a predetermined threshold value, that is, if one of them indicates that the corresponding possible repetition period exhibits a smaller coding distortion, the repetition period coder 28 selects the possible repetition period of the driving excitation source providing the evaluation value D. In contrast, when the difference between the two calculated evaluation values is less than the predetermined threshold value, the repetition period coder 28 selects one possible repetition period of the driving excitation source that is the closest to an estimate of the pitch-period of an input speech which was separately made through analysis. In either case, the repetition period coder 28 furnishes selection information coded in one bit indicating the selection result, and excitation source location code indicating the locations of the plurality of excitation sources from the driving excitation source coder 27, and polarity code indicating the polarities of the plurality of excitation sources as driving excitation source code to a multiplexer 7 as shown in Fig. 14. The repetition period coder 28 also furnishes a time-series vector associated with the driving

excitation source code, as a driving excitation source, to a gain coding unit 6 as shown in Fig. 14.

The description will be directed to the operation of the speed decoding apparatus with reference to Fig. 2. In the speech decoding apparatus having the same overall structure as shown in Fig. 15, a separator 9 separates speech code 8 output from the speech coding apparatus into linear prediction coefficient code, adaptive excitation source code, driving excitation source code, and gain code. The separator 9 then delivers the linear prediction coefficient code to a linear prediction coefficient decoding unit 10, the adaptive excitation source code to an adaptive excitation source decoder 11, the driving excitation source code to the driving excitation source decoding unit 12, and the gain code to a gain decoding unit 13. The adaptive excitation source decoding unit 11, as shown in Fig. 15, of the first embodiment converts the adaptive excitation source code to the repetition period of the adaptive excitation source and furnishes it to the driving excitation source decoding unit 12. In other words, the repetition period of the adaptive excitation source from the adaptive excitation source decoding unit 11 is delivered to the repetition period pre-selecting unit 23 of Fig. 2. The selection information included in the driving excitation source code separated by the separator 9 is furnished to the repetition period decoder 29, and the excitation source location code and polarity code included in the driving excitation source code is furnished to the driving excitation source decoder 30.

The repetition period pre-selecting unit 23 of the speech decoding apparatus has the same structure as the repetition period pre-selecting unit as shown in Fig. 1 disposed within

the speech coding apparatus. The pre-selecting unit 26 pre-selects two possible repetition periods of the driving excitation source from a plurality of possible repetition periods of the driving excitation source obtained by multiplying the input repetition period of the adaptive excitation source by a plurality of constant numbers, according to comparison results from the comparator 25, and furnishes the pre-selected two candidates for the repetition period of the driving excitation source to the repetition period decoder 29.

10 The repetition period decoder 29 selects one of the pre-selected two possible repetition periods of the driving excitation source from the pre-selecting unit 26 according to the input selection information. The repetition period decoder 29 then delivers the finally-selected possible repetition period of the driving excitation source as the repetition period of the driving excitation source to the driving excitation source decoder 30. Like the prior art driving excitation source decoding unit mentioned above, the driving excitation source decoder 30 places a plurality of fixed waveforms or pulses at a plurality of locations defined by the excitation source location code, respectively, and performs a pitch-filtering process on the plurality of fixed waveforms or pulses based on the repetition period of the driving excitation source so as to generate a series of pitch-cycles each of which includes the plurality of fixed waveforms or pulses. The driving excitation source decoder 30 then outputs the time-series vector associated with the driving excitation source code as a driving excitation source.

Referring next to Figs. 3 and 4, there are illustrated diagrams for explaining a relationship between the signal to

be coded and the pitch-filtered driving excitation source locations, i.e., the locations of pulses (or fixed waveforms) placed in each pitch-cycle of the driving excitation source, in the speech coding apparatus and the speech decoding apparatus according to the first embodiment of the present invention, respectively. The signal to be coded as shown in Fig. 3 is the same as that as shown in Fig. 18, and the signal to be coded as shown in Fig. 4 is the same as that as shown in Fig. 19. Fig. 3 shows the case where the repetition period of the adaptive excitation source is approximately twice as large as the pitch-period of the input speech. Fig. 4 shows the case where the repetition period of the adaptive excitation source is approximately one-half the pitch-period of the input speech.

In the case of Fig. 3, since the repetition period of the adaptive excitation source is equal to or greater than 40 when the pitch-period of the input speech is equal to or greater than 20, the pre-selecting unit 26 pre-selects two values one-half and equal to the repetition period of the adaptive excitation source in most cases. When the difference between the evaluation values calculated during coding for the two pre-selected possible repetition periods of the driving excitation source is less than the predetermined threshold value, the repetition period decoder 29 then selects the one one-half the repetition period of the adaptive excitation source that is closer to an estimate of the pitch-period of the input speech which was separately obtained through analysis in advance. In this case, ideal pitch-filtered excitation source locations can be obtained as shown in Fig. 3. The estimate of the pitch-period has a higher probability of being proper than the repetition period of the adaptive excitation source.

In the case of Fig. 4, since the repetition period of the adaptive excitation source is less than 40 when the pitch-period of the input speech is less than 80, the pre-selecting unit 26 selects two values equal to and twice as large as the repetition period of the adaptive excitation source in most cases. When the difference between the evaluation values calculated during coding for the two selected repetition periods of the driving excitation source is less than the predetermined threshold value, the repetition period decoder 29 then selects the one twice as large as the repetition period of the adaptive excitation source which is closer to the estimate of the pitch-period of the input speech which was separately obtained through analysis in advance. In this case, ideal periodic excitation source locations can be obtained as shown in Fig. 4.

Numerous variants may be made in the exemplary embodiment shown. As previously mentioned, an algebraic excitation source represented with the locations and polarities of a number of fixed waveforms or pulses, can be used when coding the driving excitation source and when decoding the driving excitation source code, and the present invention is, however, not limited to the structure in which the algebraic excitation source is used. The present invention can be applied to a CELP speech coding apparatus and a CELP speech decoding apparatus using a learning excitation source code book, a random excitation source code book, or the like.

Instead of the use of an estimate of the pitch-period which was separately obtained in advance, the repetition period coder 28 can select one possible repetition period of the driving excitation source that minimizes the coding distortion,

i.e., maximizes the evaluation value D. As an alternative, a value obtained by averaging the repetition periods of the adaptive excitation source obtained for a few past frames can be used instead of the pitch-period.

5        Instead of the linear prediction coefficient, another spectral parameter, such as a line spectrum pair (LSP) widely used, can be used.

10        Instead of multiplying the repetition period of the adaptive excitation source by all constant numbers located within the constant number table 24, the repetition period pre-selecting unit 23 can select two constant numbers from the constant number table 26 and, after that, multiply the repetition period of the adaptive excitation source by the two selected constant numbers, respectively, to generate two  
15        possible repetition periods of the driving excitation source. In another variant, 1 can be eliminated from the constant number table 24, and the repetition period of the adaptive excitation source can be delivered directly to the pre-selecting unit 26. Although the performance improvement is reduced, the comparator  
20        25 and the pre-selecting unit 26 can be eliminated in a case where the constant number table 25 includes  $1/2$  and 1 only.

25        As previously mentioned, in accordance with the first embodiment of the present invention, the speech coding apparatus generates a plurality of candidates for the repetition period of the driving excitation source by multiplying the repetition period of the adaptive excitation source by a plurality of constant numbers, respectively, pre-selects a predetermined number of candidates from all the candidates generated, searches for excitation source code that  
30        minimizes a coding distortion for each of the predetermined

number of candidates for the repetition period of the driving excitation source, and selects one candidate from the predetermined number of candidates according to comparison results obtained by comparing coding distortions provided for the predetermined number of candidates with a predetermined threshold value, respectively. Accordingly, the speech coding apparatus can perform a pitch-filtering process so as to generate a pitch-filtered driving excitation source using the repetition period having a high probability of being the closest to the pitch-period of the input speech even when the pitch-period of the input speech is different from the repetition period of the adaptive excitation source, thereby reducing the probability of occurrence of instability in the synthesized speech. The speech coding apparatus of the present embodiment can generate high-quality speech code.

The repetition period pre-selecting unit pre-selects two candidates or possible repetition periods of the driving excitation source, and the repetition period coding unit encodes the selection information in one bit. Accordingly, the speech coding apparatus of the present embodiment can generate high-quality speech code only with a minimum additional amount of information.

In addition, the repetition period pre-selecting unit compares the repetition period of the adaptive excitation source with a predetermined threshold value and pre-selects a predetermined number of candidates for the repetition period of the driving excitation source from all candidates according to the comparison result. Accordingly, the repetition period pre-selecting unit can reject one or more candidates for the repetition period of the driving excitation source having a



lower probability of being the closest to the pitch-period of the input speech, thus eliminating driving excitation source coding processes for the rejected candidates that don't need evaluations and reducing the required amount of the selection information to be coded. Accordingly, the speech coding apparatus of the present embodiment can generate high-quality speech code only with a minimum additional amount of operations and a minimum additional amount of information.

Furthermore, since the plurality of constant numbers by which the repetition period of the adaptive excitation source is multiplied in the repetition period pre-selecting process includes  $1/2$  and  $1$ , a number of candidates for the repetition period of the driving excitation source including the one that is the closest to the pitch-period of the input speech can be selected with a high probability while those choices are few. Accordingly, the speech coding apparatus of the present embodiment can generate high-quality speech code only with a minimum additional amount of operations and a minimum additional amount of information.

As previously mentioned, in accordance with the first embodiment of the present invention, the speech decoding apparatus generates a plurality of candidates for the repetition period of the driving excitation source by multiplying the repetition period of the adaptive excitation source by a plurality of constant numbers, pre-selects a predetermined number of candidates from all the candidates generated, further selects one candidate as the repetition period of the driving excitation source from the predetermined number of candidates pre-selected according to the selection information located within the speech code, the selection

information indicating the selection of one possible repetition period of the driving excitation source made during coding, and decodes the driving excitation source code using the repetition period of the driving excitation source to reconstruct a driving  
5 excitation source. Accordingly, the speech decoding apparatus can generate a driving excitation source that is a series of pitch-cycles using the repetition period having a high probability of being the closest to the pitch-period of the input speech even when the pitch-period of the input speech code  
10 is different from the repetition period of the adaptive excitation source, thereby reducing the probability of occurrence of instability in the synthesized speech. The speech decoding apparatus of the present embodiment can reconstruct a high-quality speech.

15 The repetition period pre-selecting unit pre-selects two candidates or possible repetition periods of the driving excitation source, and the repetition period decoding unit decodes the selection information coded in one bit and indicating the selection of one possible repetition period of  
20 the driving excitation source made during coding. Accordingly, the speech decoding apparatus of the present embodiment can generate a high-quality speech only with a minimum additional amount of information.

In addition, the repetition period pre-selecting unit  
25 compares the repetition period of the adaptive excitation source with a predetermined threshold value and pre-selects a predetermined number of candidates for the repetition period of the driving excitation source from all candidates according to the comparison result. Accordingly, the repetition period  
30 pre-selecting unit can reject one or more candidates for the

repetition period of the driving excitation source having a low probability of being the closest to the pitch-period of the input speech code, thus reducing the required amount of the selection information by one or more bits required for the rejected candidates for the repetition period of the driving excitation source, which don't need evaluations. Accordingly, the speech decoding apparatus of the present embodiment can reconstruct a high-quality speech only with a minimum additional amount of operations and a minimum additional amount of information.

Furthermore, since the plurality of constant numbers by which the repetition period of the adaptive excitation source is multiplied in the repetition period pre-selecting process includes  $1/2$  and  $1$ , a number of candidates for the repetition period of the driving excitation source including the one that is the closest to the pitch-period of the input speech code can be selected with a high probability while those choices are few. Accordingly, the speech decoding apparatus of the present embodiment can generate a high-quality speech only with a minimum additional amount of operations and a minimum additional amount of information.

#### Embodiment 2

Referring next to Fig. 5, there is illustrated a block diagram of a driving excitation source coding unit of a speech coding apparatus according to a second embodiment of the present invention. The overall structure of the speech coding apparatus of this embodiment is the same as that of the aforementioned first embodiment as shown in Fig. 14. In Fig. 5, reference numeral 31 denotes a repetition period pre-

selecting unit, and numeral 33 denotes an adaptive excitation source code book contained in an adaptive excitation source coding unit 4. The repetition period pre-selecting unit 31 includes a constant number table 32, an adaptive excitation source generating unit 34, a distance calculating unit 35, and a pre-selecting unit 36.

The driving excitation source coding unit 5 of the speech coding apparatus of the second embodiment includes a driving excitation source coder 27 that operates in the same way that the prior art driving excitation source coding unit as mentioned above, and the additional repetition period pre-selecting unit 31 and the repetition period coder 28 disposed in the front and back of the driving excitation source coder 27.

Fig. 6 is a block diagram showing the structure of a driving excitation source decoding unit of a speech decoding apparatus according to the second embodiment of the present invention. The overall structure of the speech decoding apparatus is the same as that of the aforementioned first embodiment as shown in Fig. 15. In Fig. 6, reference numeral 33 denotes an adaptive excitation source code book stored in an adaptive excitation source decoding unit 11.

The driving excitation source decoding unit 12 of the speech coding apparatus of the second embodiment includes a driving excitation source decoder 30 that operates in the same way that the prior art driving excitation source decoding unit as mentioned above, and the additional repetition period pre-selecting unit 31 and the repetition period decoder 29 disposed in the front of the driving excitation source decoder 30.

Next, a description will be made as to the operation of

the speech coding apparatus with reference to Fig. 5. Like the first embodiment, the adaptive excitation source coding unit 4 delivers the repetition period of the adaptive excitation source to the repetition period pre-selecting unit 31. A signal to be coded from the adaptive excitation source coding unit 4 and a quantized linear prediction coefficient from a linear prediction coefficient coding unit 3 are input to the driving excitation source coder 27.

The constant number table 32 of the repetition period pre-selecting unit 31 stores four constant numbers:  $1/3$ ,  $1/2$ , 1, and 2. The input repetition period of the driving excitation source is multiplied by the four constant numbers, respectively, and the four multiplication results are furnished as possible repetition periods of the driving excitation source to the adaptive excitation source generating unit 34 and the pre-selecting unit 36.

The adaptive excitation source generating unit 34 generates four other adaptive excitation sources of different repetition periods which are equal to the four possible repetition periods of the driving excitation source, respectively, using a past excitation source stored in the adaptive excitation source code book 33, and furnishes the four other adaptive excitation sources generated to the distance calculating unit 35. The adaptive excitation source generating unit 34 can eliminate the generation of one possible repetition period equal to the repetition period of the adaptive excitation source input to the repetition period pre-selecting unit 31 because the adaptive excitation source coding unit 4 has already generated the adaptive excitation source of the same repetition period.

When some of the four possible repetition periods of the driving excitation source are too large or too small and therefore they are not suitable for the pitch-period, there is a possibility that adaptive excitation source code book cannot support for the generation of the four adaptive excitation sources. To avoid such a possibility, the adaptive excitation source generating unit 34 prevents one or more possible repetition periods of the driving excitation source not suitable for the pitch-period from being selected in the pre-selecting process by furnishing a zero signal or the like as each of one or more adaptive excitation sources associated with the one or more possible repetition periods of driving excitation source.

The distance calculating unit 35 calculates a distance between the third other adaptive excitation source having the same repetition period as the adaptive excitation source applied to the repetition period pre-selecting unit 31 (i.e., the adaptive excitation source output from the adaptive excitation source coding unit 4 of Fig. 14) and each of the first, second, and fourth other adaptive excitation sources having repetition periods one-third, one-half, and twice that of the input adaptive excitation source. The distance calculating unit 35 then furnishes the calculated distances to the pre-selecting unit 36.

The pre-selecting unit 36 first compares the distance between the third other adaptive excitation source and the first other adaptive excitation source having a repetition period one-third that of the third adaptive excitation source with the distance between the third other adaptive excitation source and the second other adaptive excitation source having a repetition

period one-half that of the third adaptive excitation source, and pre-selects a shorter one of them. Then the pre-selecting unit 36 further compares the selected shorter distance with the product of an averaged magnitude of the plurality of other adaptive excitation sources and a certain constant number, and pre-selects the repetition period of the other adaptive excitation source providing the shorter distance, i.e., the repetition period being one-third or one-half that of the adaptive excitation source input from the adaptive excitation source coding unit 4, and the repetition period equal to that of the adaptive excitation source input from the adaptive excitation source coding unit 4 as two possible repetition periods of the driving excitation source when the selected shorter distance is less than the product of the averaged magnitude and the constant number. Otherwise, the pre-selecting unit 36 further compares the selected shorter distance with the distance between the third other adaptive excitation source and the fourth other adaptive excitation source having a repetition period twice that of the third adaptive excitation source, and pre-selects the repetition period of the adaptive excitation source providing a shorter one of those distances and the repetition period equal to that of the adaptive excitation source input from the adaptive excitation source coding unit 4 as two possible repetition periods of the driving excitation source. It is preferable that a positive value less than 1, e.g., about 0.1 is used as the constant number.

Like the prior art driving excitation source coding unit as shown in Fig. 17, the driving excitation source coder 27 can code an algebraic excitation source using the two possible

repetition periods of the driving excitation source pre-selected by the pre-selecting unit, the quantized linear prediction coefficient, and the signal to be coded. The present invention differs from the prior art in that each of the two possible repetition periods of the driving excitation source is obtained by multiplying that of the adaptive excitation source input from the adaptive excitation source coding unit 4 by a constant number. The driving excitation source coder 27 searches for driving excitation source code that minimizes the coding distortion for each of the two possible repetition periods of the driving excitation source, and provides the locations and polarities of a plurality of excitation sources, and an evaluation value D associated with the coding distortion according to the equation (1) described above.

The repetition period coder 28 compares the respective evaluation values D for the two possible repetition periods of the driving excitation source from the driving excitation source coder 27. If the difference between them is equal to or greater than a predetermined threshold value, that is, if one of them indicates that the corresponding possible repetition period exhibits a smaller coding distortion, the repetition period coder 28 selects the possible repetition period of the driving excitation source providing the evaluation value D. In contrast, when the difference between the two calculated evaluation values is less than the predetermined threshold value, the repetition period coder 28 selects one possible repetition period of the driving excitation source that is the closest to the pitch-period obtained through analysis (i.e., an estimation result of the pitch-period of the input speech). In either case, the



repetition period coder 28 furnishes select information coded in one bit indicating the selection result, excitation source location indicating the locations of the plurality of excitation sources, and polarity code indicating the polarities of the plurality of excitation sources as driving excitation source code to a multiplexer 7 as shown in Fig. 14.

The description will be directed to the operation of the speed decoding apparatus with reference to Fig. 6. Like the first embodiment mentioned above, the repetition period of the adaptive excitation source output from the adaptive excitation source decoding unit 11 is delivered to the repetition period pre-selecting unit 31. The selection information included in the driving excitation source code separated by a separator 9 is furnished to the repetition period decoder 29, and the excitation source location code and polarity code included in the driving excitation source code are furnished to the driving excitation source decoder 30.

The repetition period pre-selecting unit 31 of the speech decoding apparatus has the same structure as the repetition period pre-selecting unit as shown in Fig. 5 disposed within the speech coding apparatus. The pre-selecting unit 21 selects two possible repetition periods of the driving excitation source from a plurality of possible repetition periods of the driving excitation source obtained by multiplying the input repetition period of the driving excitation source by a plurality of constant numbers, and furnishes the selected two possible repetition periods to the repetition period decoder 29. The repetition period decoder 29 selects one of the selected two possible repetition periods of the driving excitation source from the pre-selecting unit 26 according to

the input selection information. The repetition period decoder 29 then delivers the finally-selected possible repetition period of the driving excitation source as the repetition period of the driving excitation source to the driving excitation source decoder 30. Like the prior art driving excitation source decoding unit mentioned above, the driving excitation source decoder 30 places a plurality of fixed waveforms or pulses at respective locations defined by the excitation source location code and performs a pitch-filtering process on them placed at the locations based on the repetition period of the driving excitation source. The driving excitation source decoder 30 also delivers a time-series vector associated with the driving excitation source code as the driving excitation source.

Figs. 7, 8, and 9 are diagrams for explaining the four other adaptive excitation sources generated by the adaptive excitation source generating unit 34 disposed within the speech coding apparatus and the speech decoding apparatus in accordance with the second embodiment of the present invention.

Fig. 7 shows the case where the repetition period of the adaptive excitation source input to the repetition period pre-selecting unit is equal to the pitch-period of the input speech. Fig. 8 shows the case where the repetition period of the input adaptive excitation source is twice the pitch-period of the input speech. Fig. 9 shows the case where the repetition period of the input adaptive excitation source is three times the pitch-period of the input speech.

When the repetition period of the input adaptive excitation source is equal to the pitch-period of the input speech, the third and fourth other adaptive excitation sources

generated with repetition periods obtained by multiplying the repetition period of the input adaptive excitation source by 1 and 2 can be selected because the distance between the first other adaptive excitation source and the third other adaptive excitation source, i.e., the original adaptive excitation source input to the repetition period pre-selecting unit (i.e., the uppermost signal of the figure) and the distance between the second other adaptive excitation source and the original adaptive excitation source are relatively long, as can be seen from Fig. 7.

When the repetition period of the input adaptive excitation source is twice the pitch-period of the input speech, the second and third other adaptive excitation sources generated with repetition periods obtained by multiplying the repetition period of the input adaptive excitation source by  $1/2$  and 1 can be selected because the distance between the second other adaptive excitation source and the original adaptive excitation source input to the repetition period pre-selecting unit (i.e., the uppermost signal of the figure) is relatively short, as can be seen from Fig. 8.

When the repetition period of the input adaptive excitation source is third times the pitch-period of the input speech, the first and third other adaptive excitation sources generated with repetition periods obtained by multiplying the repetition period of the input adaptive excitation source by  $1/3$  and 1 can be selected because the distance between the first other adaptive excitation source and the original adaptive excitation source input to the repetition period pre-selecting unit (i.e., the uppermost signal of the figure) is relatively short, as can be seen from Fig. 9.

Numerous variants may be made in the exemplary embodiment shown. As previously mentioned, the algebraic excitation source represented with the locations and polarities of a number of fixed waveforms or pulses can be used when coding and decoding the driving excitation source, and the present invention is, however, not limited to the structure in which the algebraic excitation source is used. The present invention can be applied to a CELP speech coding apparatus and CELP speech decoding apparatus using learning excitation source code book, a random excitation source code book, or the like.

Instead of the use of the pitch period of the input speech which was separately obtained in advance, the repetition period coder 28 can select one possible repetition period of the driving excitation source that minimizes the coding distortion, i.e., maximizes the evaluation value D. As an alternative, a value obtained by averaging the repetition periods of the adaptive excitation source obtained for a few previous frames can be used instead of the pitch-period of the input speech.

Instead of the linear prediction coefficient, another spectrum parameter, such as a line spectrum pair or LSP widely used, can be used.

In a variant, 1 can be eliminated from the constant number table 32, and the repetition period of the adaptive excitation source can be delivered directly to the pre-selecting unit 36. Even in this case, the pre-selecting unit 36 can work in the same way. Although the performance improvement is reduced, the constant number table 32 can include  $1/2$ , 1, and 2 only.

As previously mentioned, in accordance with the second embodiment of the present invention, the speech coding apparatus generates a plurality of candidates for the

repetition period of a driving excitation source by multiplying the repetition period of an adaptive excitation source by a plurality of constant numbers, generates a plurality of other adaptive excitation sources having repetition periods

5 respectively equal to the plurality of possible repetition periods of the driving excitation source, and selects a predetermined number of candidates from all the candidates generated according to distances between any two of the plurality of other adaptive excitation sources. Accordingly,

10 the speech coding apparatus can perform a pitch-filtering process of generating a pitch-filtered driving excitation source using the repetition period having a high probability of being the closest to the pitch-period of an input speech even when the pitch-period of the input speech is different from the

15 repetition period of the original adaptive excitation source, thereby reducing the probability of occurrence of instability in the synthesized speech. The speech coding apparatus of the present embodiment can generate high-quality speech code.

The repetition period pre-selecting unit pre-selects two

20 candidates or possible repetition periods of the driving excitation source, and the repetition period coding unit encodes the selection information in one bit. Accordingly, the speech coding apparatus of the present embodiment can generate high-quality speech code only with a minimum additional amount

25 of information.

In addition, the repetition period pre-selecting unit 31 generates a plurality of other adaptive excitation sources having repetition periods respectively equal to the plurality of possible repetition periods of the driving excitation source,

30 and selects a predetermined number of candidates from all the

candidates generated according to distances between any two of the plurality of other adaptive excitation sources.

Accordingly, the repetition period pre-selecting unit can reject one or more candidates for the repetition period of the driving excitation source having a low probability of being the closest to the pitch-period of the input speech, thus eliminating driving excitation source coding processes for the rejected candidates that don't need evaluations and reducing the required amount of the selection information. Accordingly, the speech coding apparatus of the present embodiment can generate high-quality speech code only with a minimum additional amount of arithmetic operations and a minimum additional amount of information.

Furthermore, since the plurality of constant numbers by which the repetition period of the original adaptive excitation source is multiplied in the repetition period pre-selecting process includes  $1/2$  and  $1$ , a number of candidates for the repetition period of the driving excitation source including the one that is the closest to the pitch-period of the input speech can be selected with a high probability while those choices are few. Accordingly, the speech coding apparatus of the present embodiment can generate high-quality speech code only with a minimum additional amount of arithmetic operations and a minimum additional amount of information.

As previously mentioned, in accordance with the second embodiment of the present invention, the speech decoding apparatus generates a plurality of candidates for the repetition period of a driving excitation source by multiplying the repetition period of an original adaptive excitation source by a plurality of constant numbers, pre-selects a predetermined

number of candidates from all the candidates generated, further selects one candidate as the repetition period of the driving excitation source from the predetermined number of candidates pre-selected according to the selection information located within input speech code, the selection information indicating the selection of one possible repetition period of the driving excitation source made during coding, and decodes the driving excitation source code using the repetition period of the driving excitation source to reconstruct the driving excitation source. Accordingly, the speech decoding apparatus can perform a pitch-filtering process so as to generate a pitch-filtered driving excitation source using the repetition period having a high probability of being the closest to the pitch-period of the input speech even when the pitch-period of the input speech code is different from the repetition period of the original adaptive excitation source, thereby reducing the probability of occurrence of unstability in the synthesized speech. The speech decoding apparatus of the present embodiment can generate a high-quality speech.

The repetition period pre-selecting unit pre-selects two candidates or possible repetition periods of the driving excitation source, and the repetition period decoding unit decodes the selection information coded in one bit.

Accordingly, the speech decoding apparatus of the present embodiment can reconstruct a high-quality speech only with a minimum additional amount of information.

In addition, the repetition period pre-selecting unit generates a plurality of other adaptive excitation sources having repetition periods respectively equal to the plurality of possible repetition periods of the driving excitation source,

and selects a predetermined number of candidates from all the candidates generated according to distances between any two of the plurality of other adaptive excitation sources.

Accordingly, the repetition period pre-selecting unit can  
 5 reject one or more candidates for the repetition period of the driving excitation source having a low probability of being the closest to the pitch-period of the input speech code, thus eliminating driving excitation source coding processes for the rejected candidates that don't need evaluations and reducing  
 10 the required amount of the selection information. Accordingly, the speech decoding apparatus of the present embodiment can generate a high-quality speech only with a minimum additional amount of arithmetic operations and a minimum additional amount of information.

15 Furthermore, since the plurality of constant numbers by which the repetition period of the original adaptive excitation source is multiplied in the repetition period pre-selecting process includes  $1/2$  and  $1$ , a number of candidates for the repetition period of the driving excitation source including  
 20 the one that is the closest to the pitch-period of the input speech code can be selected with a high probability while those choices are few. Accordingly, the speech decoding apparatus of the present embodiment can reconstruct a high-quality speech only with a minimum additional amount of arithmetic operations  
 25 and a minimum additional amount of information.

### Embodiment 3

Referring next to Fig. 10, there is illustrated a block diagram showing the structure of a driving excitation source  
 30 coding unit 5 and a perceptual weighting control unit 37



disposed within a speech coding apparatus in accordance with a third embodiment of the present invention. The overall structure of the speech coding apparatus of this embodiment thus involves the additional perceptual weighting control unit 37  
5 connected to the driving excitation source coding unit 5 in addition to the structure as shown in Fig. 14. The perceptual weighting control unit 37 includes a comparator 38 and a strength control unit 39. The driving excitation source coding unit 5 has the same structure as the conventional driving  
10 excitation source coding unit as shown in Fig. 17, with the exception that a perceptual weighting filter coefficient calculating unit 16 is controlled by the perceptual weighting control unit 37.

In operation, a linear prediction coefficient coding unit  
15 3, as shown in Fig. 14, of the speech coding apparatus delivers a quantized linear prediction coefficient to the perceptual weighting filter coefficient calculating unit 16 and a basic response generating unit 18 disposed within the driving  
20 excitation source coding unit 5. An adaptive excitation source coding unit 4 converts adaptive excitation source code into a repetition period of an adaptive excitation source and then furnishes the repetition period of the adaptive excitation  
25 source to the basic response generating unit 18 of the driving excitation source coding unit 5 and the comparator 38 of the perceptual weighting control unit 37. The adaptive excitation source coding unit 4 also delivers either an input speech 1 or a signal obtained by subtracting a synthesized speech generated based on the adaptive excitation source from the input speech  
30 1, as a signal to be coded, to a perceptual weighting filter 17.

The comparator 38 of the perceptual weighting control unit 37 compares the input repetition period of the adaptive excitation source with a predetermined threshold value and furnishes the comparison result to the strength control unit 39. The predetermined threshold value can be about 40 which can substantially separate the distribution of pitch-periods into a male-speech region and a female-speech region.

The strength control unit 39 determines the strength coefficient to control an enhanced strength for the perceptual weighting filter 17 and another perceptual weighting filter 19 according to the comparison result from the comparator 38, and furnishes the determined strength coefficient to the perceptual weighting filter coefficient calculating unit 16 of the driving excitation source coding unit 5. When the comparison result from the comparator 38 indicates that the repetition period of the adaptive excitation source is equal to or greater than the predetermined threshold value, the strength control unit 39 determines the strength coefficient so that the perceptual weighting strength becomes lower because there is a high possibility that the speech to be coded is a male speech. In contrast, when the comparison result from the comparator 38 indicates that the repetition period of the adaptive excitation source is less than the predetermined threshold value, the strength control unit 39 determines the strength coefficient so that the perceptual weighting strength becomes higher because there is a high possibility that the speech to be coded is a female speech. A multiplier by which the linear prediction coefficient is multiplied, the linear prediction coefficient being used for calculating the perceptual weighting filter coefficient, can be used as the strength coefficient, for

example.

The perceptual weighting filter coefficient calculating unit 16 calculates the perceptual weighting filter coefficient using the quantized linear prediction coefficient and the strength coefficient, and defines the calculated perceptual weighting filter coefficient as a filter coefficient for the two perceptual weighting filters 17 and 19.

After that, the first perceptual weighting filter 17, the basis response generating unit 18, the second perceptual weighting filter 19, a pre-table calculating unit 20, a searching unit 21, and an excitation source location table 22 operate in the same way that the same components of conventional speech coding apparatuses mentioned above do, and therefore the description of the operations of those components will be omitted hereinafter.

Numerous variants may be made in the exemplary embodiment shown. It is clear that instead of determining the strength coefficient according to whether or not the repetition period of the adaptive excitation source is equal to or greater than a predetermined threshold value, the perceptual weighting control unit 37 can control the strength coefficient more finely using two or more predetermined threshold values or continuously control the strength coefficient according to the difference between the repetition period of the adaptive excitation source and a predetermined threshold value.

The present embodiment is not limited to the above-mentioned algebraic excitation source arrangement using algebraic excitation sources when coding the driving excitation source, and can be applied to a CELP speech coding apparatus using a learning excitation source code book, a random

excitation source code book, or the like.

Instead of the linear prediction coefficient, another spectrum parameter, such as a line spectrum pair or LSP widely used, can be used.

5 As previously mentioned, in accordance with the third embodiment of the present invention, the speech coding apparatus controls the perceptual weighting strength coefficient based on the repetition period of the adaptive excitation source, calculates the filter coefficient for the  
10 two perceptual weighting filters using the perceptual weighting strength coefficient, and performs a perceptual weighting process on the signal to be coded, which is used for coding the driving excitation source. Accordingly, the perceptual weighting process can be optimized for male and female speeches,  
15 and the speech coding apparatus of the third embodiment can provide high-quality speech code.

#### Embodiment 4

Referring next to Fig. 11, there is illustrated a block  
20 diagram showing the structure of a driving excitation source coding unit 5 and an additional perceptual weighting control unit 40 disposed within a speech coding apparatus in accordance with a fourth embodiment of the present invention. The overall structure of the speech coding apparatus of this embodiment thus  
25 involves the additional perceptual weighting control unit 40 connected to the driving excitation source coding unit 5 in addition to the structure as shown in Fig. 14. The perceptual weighting control unit 40 includes a comparator 38, a strength control unit 39, and an average updating unit 41. The driving  
30 excitation source coding unit 5 has the same structure as the

conventional driving excitation source coding unit as shown in Fig. 17, with the exception that a perceptual weighting filter coefficient calculating unit 16 is controlled by the perceptual weighting control unit 40.

5        Since the present embodiment differs from the above-mentioned third embodiment in that the perceptual weighting control unit 40 includes the average updating unit 41 in addition to the structure of the perceptual weighting control unit 37 of the third embodiment, the description will be mainly  
10        directed to the operation of the additional component. An adaptive excitation source coding unit 4 converts an adaptive excitation source code into a repetition period of an adaptive excitation source and then furnishes the repetition period of the adaptive excitation source to a basic response generating  
15        unit 18 of the driving excitation source coding unit 5 and the average updating unit 41 of the perceptual weighting control unit 40.

      The average updating unit 41 of the perceptual weighting control unit 40 updates an average of previously stored  
20        repetition periods of the adaptive excitation source using the input repetition period of the adaptive excitation source, and delivers the averaged repetition period to the comparator 38. There can be provided some methods of easily updating the average including an averaging method of calculating the sum  
25        of the product of the repetition period of the adaptive excitation source associated with the current frame and a constant number  $\alpha$  less than 1 and the product of the previous average and  $(1-\alpha)$ . Since the aim of obtaining the average is to precisely determine whether the input speech is a male speech  
30        or a female speech, it is preferable to limit the updating to

frames with a large adaptive excitation source gain.

The comparator 38 compares the updated average with a predetermined threshold value and furnishes the comparison result to the strength control unit 39. The strength control unit 39 determines a strength coefficient to control an enhanced strength for perceptual weighting filters 17 and 19 based on the comparison result from the comparator 38, and furnishes the determined strength coefficient to the perceptual weighting filter coefficient calculating unit 16 of the driving excitation source coding unit 5. When the comparison result from the comparator 38 indicates that the average is equal to or greater than the predetermined threshold value, the strength control unit 39 determines the strength coefficient so that the perceptual weighting strength becomes lower because there is a high possibility that the speech to be coded is a male speech. In contrast, when the comparison result from the comparator 38 indicates that the average is less than the predetermined threshold value, the strength control unit 39 determines the strength coefficient so that the perceptual weighting strength becomes higher because there is a high possibility that the speech to be coded is a female speech.

After that, the perceptual weighting filter coefficient calculating unit 16, the first perceptual weighting filter 17, the basis response generating unit 18, the second perceptual weighting filter 19, a pre-table calculating unit 20, a searching unit 21, and an excitation source location table 22 operate in the same way that the same components of conventional speech coding apparatuses as shown in Fig. 17 do, and therefore the description of the operations of those components will be omitted hereinafter.

Numerous variants may be made in the exemplary embodiment shown. It is clear that instead of determining the strength coefficient according to whether or not the averaged repetition period of the adaptive excitation source is equal to or greater  
5 than a predetermined threshold value, the perceptual weighting control unit 40 can control the strength coefficient more finely using two or more predetermined threshold values or continuously control the strength coefficient according to the difference between the averaged repetition period of the  
10 adaptive excitation source and a predetermined threshold value.

The present embodiment is not limited to the above-mentioned algebraic excitation source arrangement using algebraic excitation sources when coding the driving excitation source, and can be applied to a CELP speech coding apparatus  
15 using a learning excitation source code book, a random excitation source code book, or the like.

Instead of the linear prediction coefficient, another spectrum parameter, such as a line spectrum pair or LSP widely used, can be used.

20 As previously mentioned, in accordance with the fourth embodiment of the present invention, the speech coding apparatus controls the perceptual weighting strength coefficient based on the averaged repetition period of the adaptive excitation source, calculates the filter coefficient  
25 for the two perceptual weighting filters using the perceptual weighting strength coefficient, and performs a perceptual weighting process on the signal to be coded, which is used for coding the driving excitation source. Accordingly, the perceptual weighting process can be optimized for male and  
30 female speeches, and the speech coding apparatus of the fourth

embodiment can provide high-quality speech code.

Because of the use of the averaged repetition period of the adaptive excitation source, the present embodiment can prevent the perceptual weighting strength from frequently  
5 varying and hence reduce the occurrence of unstability in the speech code.

#### Embodiment 5

Referring next to Fig. 12, there is illustrated an  
10 excitation source location table 22 which is used by a driving excitation source coding unit 5 of a speech coding apparatus according to a fifth embodiment of the present invention and a driving excitation source decoding unit 12 of a speech  
decoding apparatus according to the fifth embodiment. The  
15 excitation source location table 22 of this embodiment further includes a certain magnitude for each of a plurality of excitation source numbers in addition to the same elements as the prior art excitation source location table as shown in Fig.  
16.

20 In the same excitation source location table, the fixed magnitude provided for each of the plurality of excitation source numbers depends on the number of candidates for the excitation source location provided for a corresponding  
excitation source number. In the example as shown in Fig. 12,  
25 each of the excitation source numbers starting from No. 1 to 3 includes 8 candidates for the excitation source location and the same fixed magnitude of 1.0. Since the number of candidates included in the last excitation source number, i.e., No. 4 is  
16, which is greater than the number of candidates included in  
30 any other excitation source number, a fixed magnitude of 1.2



larger than any other fixed magnitude in the same location table is provided for the excitation source number 4. In this manner, the larger the number of candidates for the excitation source location, the larger a fixed magnitude is provided.

- 5 Searching for an optimum combination of excitation source locations using the excitation source location table having the additional fixed magnitudes can be performed based on the above-mentioned equation (1). In this embodiment, C and E of the equation (1) are given by:

10

$$C = \sum_k d''(m_k) \quad (8)$$

$$E = \sum_k \sum_i \phi''(m_k, m_i) \quad (9)$$

$d''(m_k)$  and  $\phi''(m_k, m_i)$  are given by:

$$d''(m_k) = a_k d'(m_k) \quad (10)$$

$$15 \quad \phi''(m_k, m_i) = a_k a_i \phi'(m_k, m_i) \quad (11)$$

- where  $a_k$  is the magnitude of the  $k$ th pulse, which is equal to one magnitude listed in the excitation source location table of Fig. 12. Only calculating and storing  $d''(m_k)$  and  $\phi''(m_k, m_i)$  as a pre-table in advance of the calculation of the evaluation value D for all combinations of all pulse locations is thus  
20 needed before the simple summations according to the equations (8) and (9), thereby reducing the amount of arithmetic operations.

- 25 The decoding of the driving excitation source can be performed by selecting one excitation source location for each

of the plurality of excitation source numbers stored in the excitation source location table of Fig. 12 based on the excitation source location code, and for placing an excitation source, which is then multiplied by the fixed magnitude provided for each of the plurality of excitation source numbers, at a corresponding excitation source location selected for each of the plurality of excitation source numbers. When each of the plurality of excitation sources placed is not a pulse or when generating a series of pitch-cycles each includes the plurality of excitation sources, elements of the plurality of excitation sources placed overlap and all that is needed is to calculate the sum of all overlapped portions. In other words, the driving excitation source decoding process of the present embodiment includes the process of multiplying a plurality of excitation sources to be placed by respective fixed magnitudes provided for the plurality of excitation source numbers in addition to the conventional algebraic excitation source decoding process.

In a prior art decoding process in which a fixed waveform is prepared for each of the plurality of excitation source numbers, a basic response has to be calculated for each of the plurality of excitation source numbers. In contrast, in accordance with the present embodiment, only a modification of the pre-table is added as previously mentioned. In any prior art decoding process, the magnitude of each of the plurality of excitation sources is maintained constant even though the amount of location information (i.e., the number of candidates for the excitation source location) varies from excitation source number to excitation source number.

As previously mentioned, in accordance with the fifth embodiment of the present invention, the speech coding

apparatus provides a certain magnitude depending on the number of candidates for the location of each of a plurality of excitation sources for each of the plurality of excitation sources and multiplies the plurality of excitation sources placed at respective possible locations by the plurality of fixed magnitudes, respectively, by means of the driving excitation source coding unit 5. The driving excitation source coding unit 5 then generates a driving excitation source by calculating the sum of all the excitation sources placed at the respective possible locations for each of all combinations of possible locations of the plurality of excitation sources, and searches for excitation source code and polarity code associated with one driving excitation source exhibiting the smallest coding distortion between itself and the input speech, the excitation source code indicating the locations of the plurality of excitation sources placed and the polarity code indicating the polarities of the plurality of excitation sources placed. The speech coding apparatus can avoid waste concerned with the setting of the magnitudes of the plurality of excitation sources to a fixed value, and generate high-quality speech code.

Similarly, in accordance with the fifth embodiment of the present invention, the speech decoding apparatus provides a certain magnitude depending on the number of candidates for the location of each of a plurality of excitation sources for each of the plurality of excitation sources. The driving excitation source decoding unit 12 then generates a driving excitation source by calculating the sum of all the excitation sources placed at respective possible locations defined by the excitation source location coded included in the input speech

code while multiplying the plurality of excitation sources placed at the respective possible locations by the plurality of fixed magnitudes, respectively. The speech decoding apparatus can avoid waste concerned with the setting of the magnitudes of the plurality of excitation sources to a fixed value, and reconstruct a high-quality speech.

#### Embodiment 6

Referring next to Fig. 13, there is illustrated a block diagram showing the structure of a driving excitation source coding unit 5 of a speech coding apparatus in accordance with a sixth embodiment of the present invention. The overall structure of the speech coding apparatus of this embodiment is the same as that of prior art speech coding apparatuses as shown in Fig. 14. In Fig. 13, reference numeral 42 denotes a pre-table modifying unit. The speech coding apparatus of the sixth embodiment can make a perceptual weighted signal to be coded orthogonal to an adaptive excitation source using only the additional pre-table modifying unit 42.

In operation, a linear prediction coefficient coding unit 3 delivers a quantized linear prediction coefficient to both a perceptual weighting filter coefficient calculating unit 16 disposed within the driving excitation source coding unit 5 and a basic response generating unit 18. An adaptive excitation source coding unit 4 converts an adaptive excitation source code into a repetition period of an adaptive excitation source and then furnishes the repetition period of the adaptive excitation source to the basic response generating unit 18 located within the driving excitation source coding unit 5. The adaptive excitation source coding unit 4 also delivers either an input

speech 1 or a signal obtained by subtracting a synthesized speech generated based on the adaptive excitation source from the input speech 1, as a signal to be coded, to a perceptual weighting filter 17. The adaptive excitation source coding unit 4 further furnishes the adaptive excitation source to the pre-table modifying unit 42 located within the driving excitation source coding unit 5.

The perceptual weighting filter coefficient calculating unit 16 calculates a perceptual weighting filter coefficient using the quantized linear prediction coefficient and defines the calculated perceptual weighting filter coefficient as a filter coefficient for the perceptual weighting filter 17 and another perceptual weighting filter 19. The perceptual weighting filter 17 performs a filtering process on the input signal to be coded using the filter coefficient set by the perceptual weighting filter coefficient calculating unit 16.

The basic response generating unit 18 performs a pitch-filtering process on either a unit pulse or a fixed waveform using the input repetition period of the adaptive excitation source so as to generate a series of pitch-cycles each of which includes either the unit pulse or the fixed waveform. The basic response generating unit 18 then generates a synthesized speech by allowing the generated signal as an excitation source to pass through a synthesis filter constructed using the quantized linear prediction coefficient, and furnishes the synthesized speech as a basic response to the perceptual weighting filter 19. The perceptual weighting filter 19 performs a filtering process on the input basic response using the filter coefficient set by the perceptual weighting filter coefficient calculating unit 16.

The pre-table calculating unit 20 calculates a correlation  $d(x)$  between the perceptual weighed signal to be coded from the perceptual weighting filter 17 and each of the plurality of perceptual weighed basic responses from the perceptual weighting filter 19, i.e., each of a plurality of perceptual weighed synthesized speeches respectively generated based on a plurality of temporary driving excitation sources, which are signals obtained by placing a predetermined excitation source at all possible excitation source locations, respectively. The pre-table calculating unit 20 also calculates a cross-correlation  $\phi(x,y)$  between any two of the plurality of perceptual weighted basic responses, i.e., any two of the plurality of synthesized speeches respectively generated based on the plurality of temporary driving excitation sources.  $d(x)$  and  $\phi(x,y)$  are stored as a pre-table.

The pre-table modifying unit 42 accepts the adaptive excitation source and the pre-table stored in the pre-table calculating unit 20 and modifies the pre-table according to the following equations (12) and (13). The pre-table modifying unit 42 then calculates  $d'(x)$  and  $\phi'(x,y)$  according to the following equations (14) and (15) and stores these parameters as a new pre-table.

$$\hat{d}(x) = d(x) - \frac{c_x c_{tgt}}{P_{acb}} \quad (12)$$

$$\hat{\phi}(x, y) = \phi(x, y) - \frac{c_x c_y}{P_{acb}} \quad (13)$$

$$d'(m_k) = |\hat{d}(m_k)| \quad (14)$$

$$\phi'(m_k, m_i) = \text{sign}[\hat{d}(m_k)] \text{sign}[\hat{d}(m_i)] \hat{\phi}(m_k, m_i) \quad (15)$$

where  $c_{tgt}$  is a correlation between the perceptual weighted signal to be coded and a perceptual weighted adaptive excitation source response (i.e., synthesized speech), i.e., a correlation between the perceptual weighted signal to be coded and a synthesized speech generated based on the perceptual weighted adaptive excitation source,  $c_x$  is a correlation between a signal created by placing the perceptual weighted basic response at the excitation source location  $x$  and the perceptual weighted adaptive excitation source response (i.e., synthesized speech), i.e., a correlation between each of the plurality of perceptual weighed synthesized speeches respectively generated based on the plurality of temporary driving excitation sources and the synthesized speech generated based on the adaptive excitation source, and  $p_{acb}$  is the power of the perceptual weighted adaptive excitation source response (i.e., synthesized speech).

The searching unit 21 sequentially reads the plurality of candidates for the excitation source location from the excitation source location table 22, and calculates the evaluation value  $D$  for each of all combinations of possible

excitation source locations using the pre-table stored in the pre-table modifying unit 42, i.e.,  $d'(x)$  and  $\phi'(x,y)$  calculated for each of all combinations of possible excitation source locations according to the equations (1), (4) and (5). The  
 5 searching unit 21 then searches for one combination of excitation source locations that maximizes the evaluation value  $D$  and furnishes excitation source location code (i.e., indexes of the excitation source location table) indicating the plurality of possible excitation source locations searched for  
 10 and polarity code indicating the polarities of the plurality of excitation sources, as driving excitation source code. The searching unit 21 generates a time-series vector associated with the driving excitation source code as a driving excitation source.

15 As previously mentioned, in accordance with the sixth embodiment of the present invention, the speech coding apparatus calculates a correlation  $c_{tgt}$  between the perceptual weighted signal to be coded and a synthesized speech generated based on the perceptual weighted adaptive excitation source,  
 20 and a correlation  $c_x$  between each of a plurality of perceptual weighed synthesized speeches respectively generated based on a plurality of temporary driving excitation sources, which are associated with all possible excitation source locations, respectively, and the synthesized speech generated based on the  
 25 adaptive excitation source, and then modifies the pre-table using these correlations. Accordingly, the speech coding apparatus can make the perceptual weighted signal to be coded orthogonal to the adaptive excitation source without increase in the amount of arithmetic operations in the searching unit  
 30 21, thereby improving the coding performance and providing





## WHAT IS CLAIMED IS:

1. A speech coding apparatus for coding an input speech on a fame-by-frame basis using an adaptive excitation source, which is generated from a past excitation source, and a driving  
5 excitation source, which is generated from the input speech and the adaptive excitation source, so as to generate speech code, said speech coding apparatus comprising:

a repetition period pre-selecting means for generating a plurality of candidates for a repetition period of the driving  
10 excitation source by multiplying a repetition period of the adaptive excitation source by a plurality of constant numbers, respectively, and for pre-selecting a predetermined number of candidates from all the candidates generated and furnishing the predetermined number of pre-selected candidates;

15 a driving excitation source coding means for providing both excitation source location information and excitation source polarity information that minimize a coding distortion, for each of the predetermined number of candidates for the repetition period of the driving excitation source, and for  
20 providing an evaluation value associated with the minimum coding distortion for each of the predetermined number of candidates; and

a repetition period coding means for comparing the evaluation values provided for the predetermined number of  
25 candidates for the repetition period of the driving excitation source from said driving excitation source coding means with one another, for selecting one candidate from the predetermined number of candidates according to a comparison result, and for furnishing selection information indicating a selection result,  
30 excitation source location code indicating excitation source

location information associated with the selected candidate for the repetition period of the driving excitation source, and polarity code indicating excitation source polarity information associated with the selected candidate.

5

2. The speech coding apparatus according to Claim 1, wherein said repetition period pre-selecting means pre-selects two candidates from all the candidates generated, and said repetition period coding means encodes the selection result in one bit so as to generate 1-bit selection information.

10

3. The speech coding apparatus according to Claim 1, wherein said repetition period pre-selecting means includes a means for comparing the repetition period of the adaptive excitation source with a predetermined threshold value, and for pre-selecting the predetermined number of candidates from all the candidates generated according to a comparison result.

15

4. The speech coding apparatus according to Claim 1, wherein said repetition period pre-selecting means includes a means for generating a plurality of other adaptive excitation sources whose respective repetition periods equal to the plurality of candidates for the repetition period of the driving excitation source, respectively, and for pre-selecting the predetermined number of candidates from all the candidates generated according to a comparison between distances among the plurality of other adaptive excitation sources generated.

20

25

5. The speech coding apparatus according to Claim 1, wherein said plurality of constant numbers, by which the

30

repetition period of the adaptive excitation source is multiplied, includes  $1/2$  and 1.

6. A speech decoding apparatus for decoding input speech  
5 code on a fame-by-frame basis using an adaptive excitation  
source, which is generated from a past excitation source, and  
a driving excitation source, which is generated from the input  
speech code and the adaptive excitation source, so as to  
reconstruct original speech, said speech decoding apparatus  
10 comprising:

a repetition period pre-selecting means for providing a  
plurality of candidates for a repetition period of the driving  
excitation source by multiplying a repetition period of the  
adaptive excitation source by a plurality of constant numbers,  
15 respectively, and for pre-selecting a predetermined number of  
candidates from all the candidates generated and furnishing the  
predetermined number of pre-selected candidates;

a repetition period decoding means for selecting one  
candidate from the predetermined number of pre-selected  
20 candidates for the repetition period of the driving excitation  
source from said repetition period pre-selecting means  
according to selection information included in said input coded  
speech and indicating the selection, and for furnishing the  
selected candidate as the repetition period of the driving  
25 excitation source; and

a driving excitation source decoding means for generating  
a time-series signal according to excitation source location  
code and excitation source polarity code included in the input  
speech code, and for generating a time-series vector that is  
30 a series of pitch-cycles, each of which includes the time-series

signal, using the repetition period of the driving excitation source from said repetition period decoding means.

7. The speech decoding apparatus according to Claim 6,  
5 wherein said repetition period pre-selecting means pre-selects two candidates from all the candidates generated, and said repetition period decoding means decodes selection information coded in one bit, which is included in the input speech code and indicates a selection of a candidate for the repetition  
10 period of the adaptive excitation source made during coding.

8. The speech decoding apparatus according to Claim 6,  
wherein said repetition period pre-selecting means includes a means for comparing the repetition period of the adaptive  
15 excitation source with a predetermined threshold value, and for pre-selecting the predetermined number of candidates from all the candidates generated according to a comparison result.

9. The speech decoding apparatus according to Claim 6,  
20 wherein said repetition period pre-selecting means includes a means for generating a plurality of other adaptive excitation sources whose respective repetition periods equal to the plurality of candidates for the repetition period of the driving excitation source, respectively, and for pre-selecting the  
25 predetermined number of candidates from all the candidates generated according to a comparison between distances among the plurality of other adaptive excitation sources generated.

10. The speech decoding apparatus according to Claim 6,  
30 wherein the plurality of constant numbers, by which the

repetition period of the adaptive excitation source is multiplied, includes  $1/2$  and  $1$ .

11. A speech coding apparatus for coding an input speech  
5 on a fame-by-frame basis using an adaptive excitation source,  
which is generated from a past excitation source, and a driving  
excitation source, which is generated from the input speech and  
the adaptive excitation source, so as to generate speech code,  
said speech coding apparatus comprising:

10 a perceptual weighting control means for determining a  
perceptual weighting strength coefficient based on a repetition  
period of the adaptive excitation source; and

a driving excitation source coding means for generating  
excitation source location code indicating information about  
15 excitation source locations and information about excitation  
source polarities based on the repetition period of the adaptive  
excitation source, the perceptual weighting strength  
coefficient determined by said perceptual weighting control  
means, and a signal to be coded such as the input speech.

20

12. The speech coding apparatus according to Claim 11,  
wherein said perceptual weighting control means determines the  
perceptual weighting strength coefficient based on an average  
of the repetition period of the current adaptive excitation  
25 source and repetition periods of previously-generated adaptive  
excitation sources.

13. A speech coding apparatus for coding an input speech  
on a fame-by-frame basis using an adaptive excitation source,  
30 which is generated from a past excitation source, and a driving

excitation source generated from the input speech and the adaptive excitation source, said driving excitation source being represented by locations and polarities of a plurality of excitation sources, so as to generate speech code, said  
5 speech coding apparatus comprising:

an excitation source location table including a plurality of selectable possible locations and a fixed magnitude determined based on the number of the plurality of possible locations for each of the plurality of excitation sources;  
10 a driving excitation source coding means for placing the plurality of excitation sources at respective possible locations while multiplying each of the plurality of excitation sources by a corresponding fixed magnitude, with reference to said excitation source location table, for generating a driving  
15 excitation source by calculating a sum of the plurality of excitation sources each of which has been multiplied by the corresponding fixed magnitude and is thus placed at one corresponding possible location, for each of all combinations of possible locations of the plurality of excitation sources,  
20 and for selecting possible locations and polarities of the plurality of excitation sources which provide a driving excitation source having a smallest coding distortion between itself and the input speech so as to generate excitation source location code and polarity code.

25

14. A speech decoding apparatus for decoding input speech code on a frame-by-frame basis using an adaptive excitation source, which is generated from a past excitation source, and a driving excitation source generated from the input speech code  
30 and the adaptive excitation source, said driving excitation

source being represented by locations and polarities of a plurality of excitation sources, so as to reconstruct original speech, said speech decoding apparatus comprising:

an excitation source location table including a plurality  
 5 of selectable possible locations and a fixed magnitude determined based on the number of the plurality of possible locations for each of the plurality of excitation sources;

a driving excitation source decoding means for selecting  
 10 respective possible locations for the plurality of excitation sources with reference to said excitation source location table based on excitation source location code included in the input speech code, for placing the plurality of excitation sources at the respective selected possible locations while multiplying  
 15 each of the plurality of excitation sources by a corresponding fixed magnitude, and for generating a driving excitation source by calculating a sum of the plurality of excitation sources each of which has been multiplied by the corresponding fixed magnitude and is thus placed at the corresponding possible location.

20

15. A speech coding apparatus for coding an input speech on a frame-by-frame basis using an adaptive excitation source, which is generated from a past excitation source, and a driving excitation source generated from the input speech and the  
 25 adaptive excitation source, said driving excitation source being represented by locations and polarities of a plurality of excitation sources, so as to generate speech code, said speech coding apparatus comprising:

a pre-table calculating means for calculating a  
 30 correlation between a signal to be coded, such as the input



speech, and each of a plurality of synthesized speeches each of which is generated based on a corresponding temporary driving excitation source that is a signal obtained by placing a predetermined excitation source at a corresponding one of all possible locations, and a cross-correlation between any two of the plurality of synthesized speeches, and for storing these calculated correlations and cross-correlations as a pre-table therein;

10 a pre-table modifying means for calculating a correlation between the signal to be coded and a synthesized speech generated based on the adaptive excitation source, and a correlation between each of the plurality of synthesized speeches generated based on the corresponding temporary driving excitation source and the synthesized speech generated based on the adaptive excitation source, and for modifying said pre-table using these calculated correlations; and

20 a searching means for determining the locations and polarities of the plurality of excitation sources using the pre-table corrected by said pre-table modifying means so as to generate excitation source location code indicating the locations of the plurality of excitation sources and excitation source polarity code indicating the polarities of the plurality of excitation sources.

## ABSTRACT OF THE DISCLOSURE

A speech coding apparatus comprises a repetition period pre-selecting unit for generating a plurality of candidates for the repetition period of a driving excitation source by

5 multiplying the repetition period of an adaptive excitation source by a plurality of constant numbers, respectively, and for pre-selecting a predetermined number of candidates from all the candidates generated. A driving excitation source coding unit provides both excitation source location information and

10 excitation source polarity information that minimize a coding distortion, for each of the predetermined number of candidates, and provides an evaluation value associated with the minimum coding distortion for each of the predetermined number of candidates. A repetition period coding unit compares the

15 evaluation values provided for the predetermined number of candidates with one another, selects one candidate from the predetermined number of candidates according to the comparison result, and furnishes selection information indicating the selection result, excitation source location code, and polarity

20 code.

FIG. 1

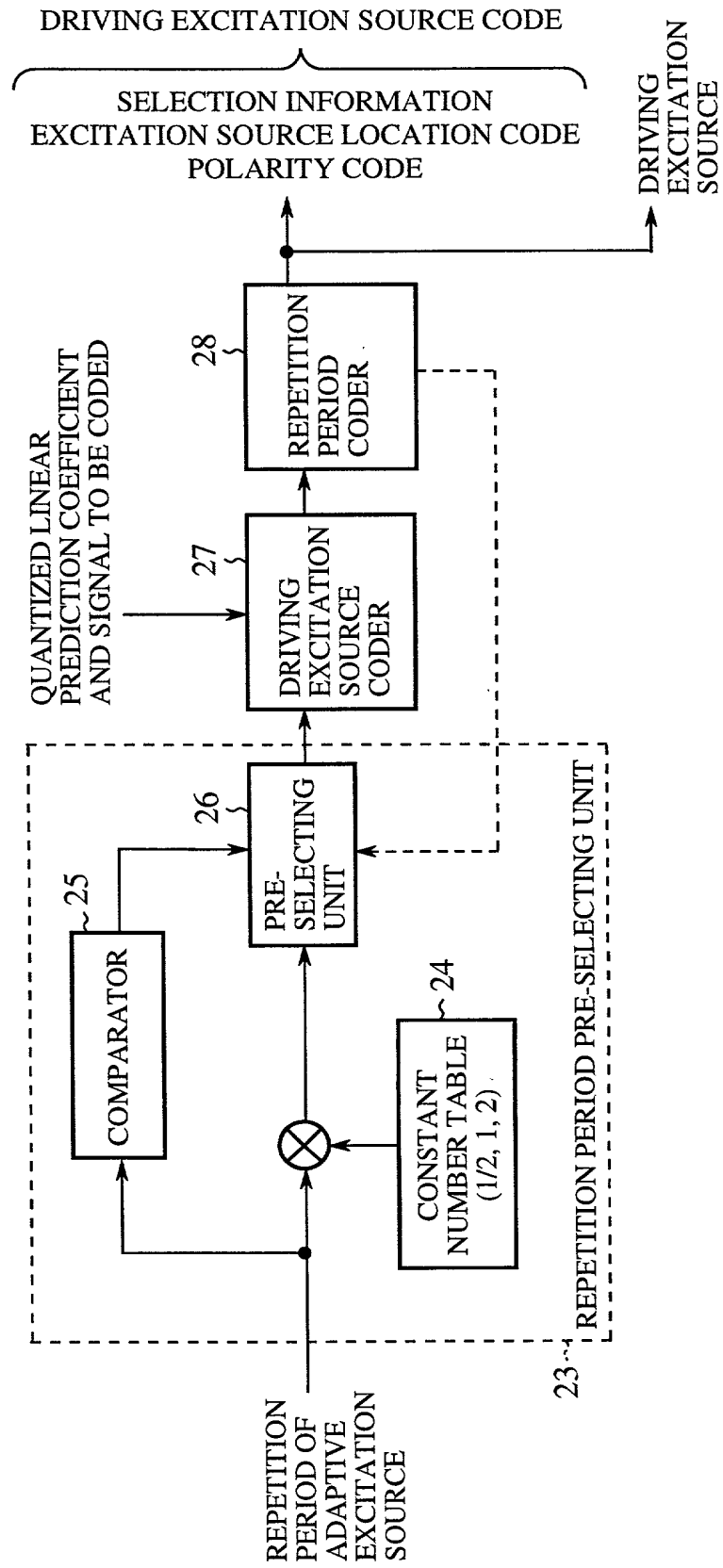


FIG.2

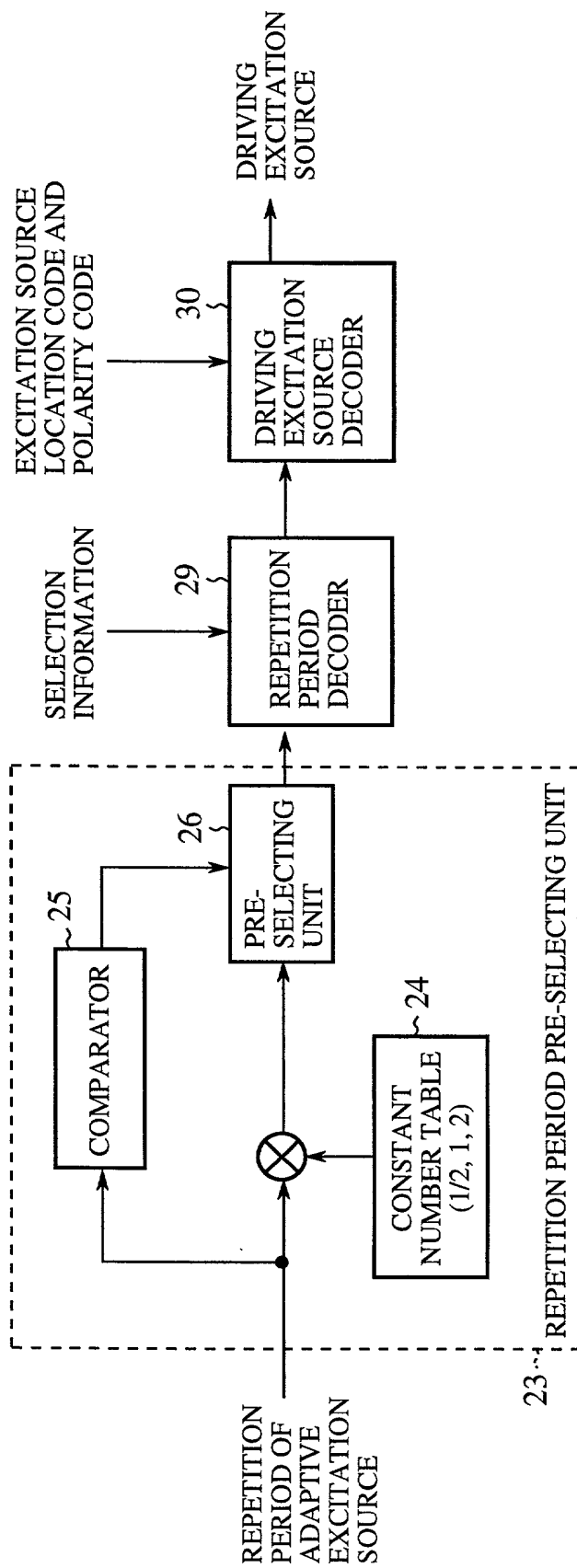


FIG.3

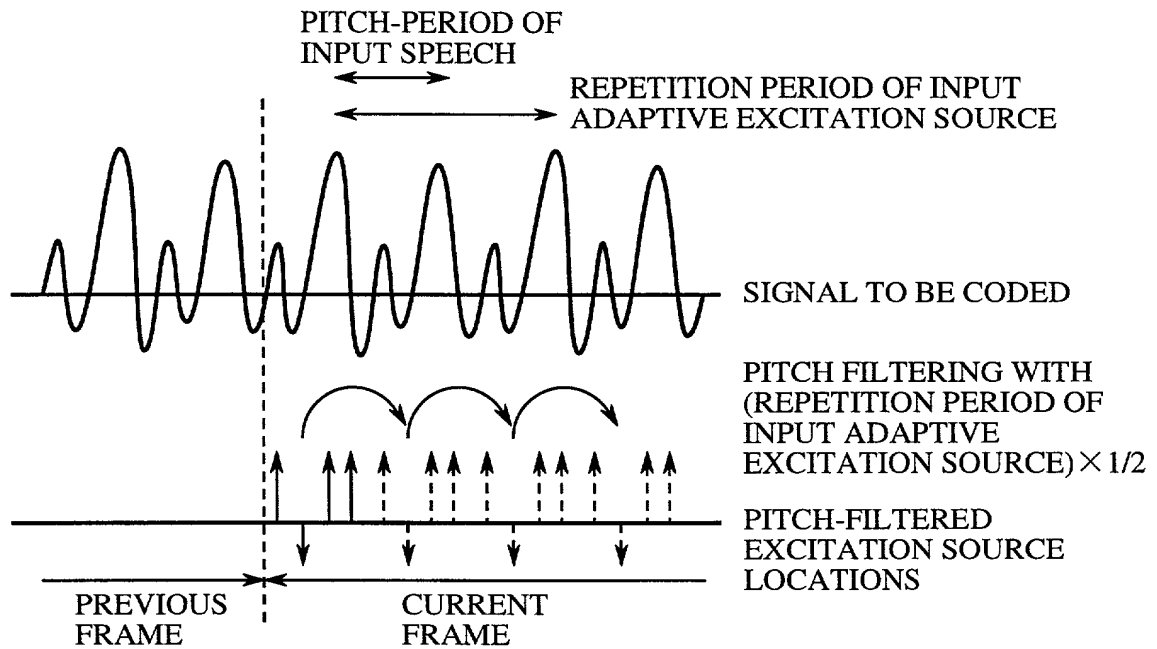


FIG.4

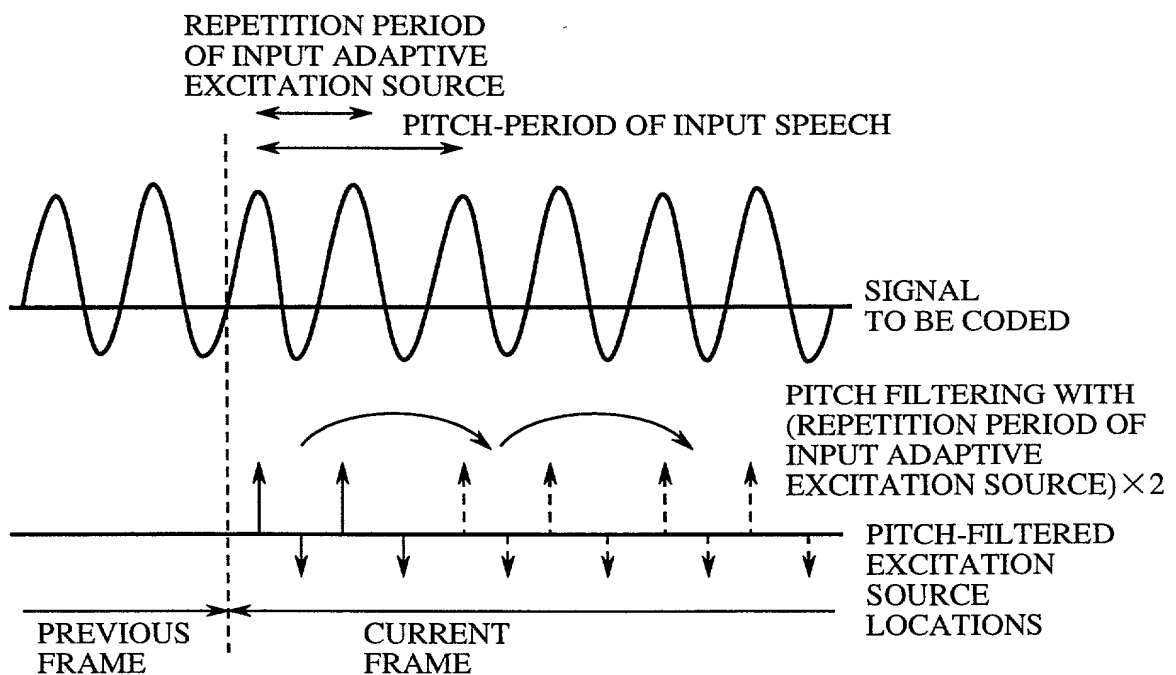


FIG.5

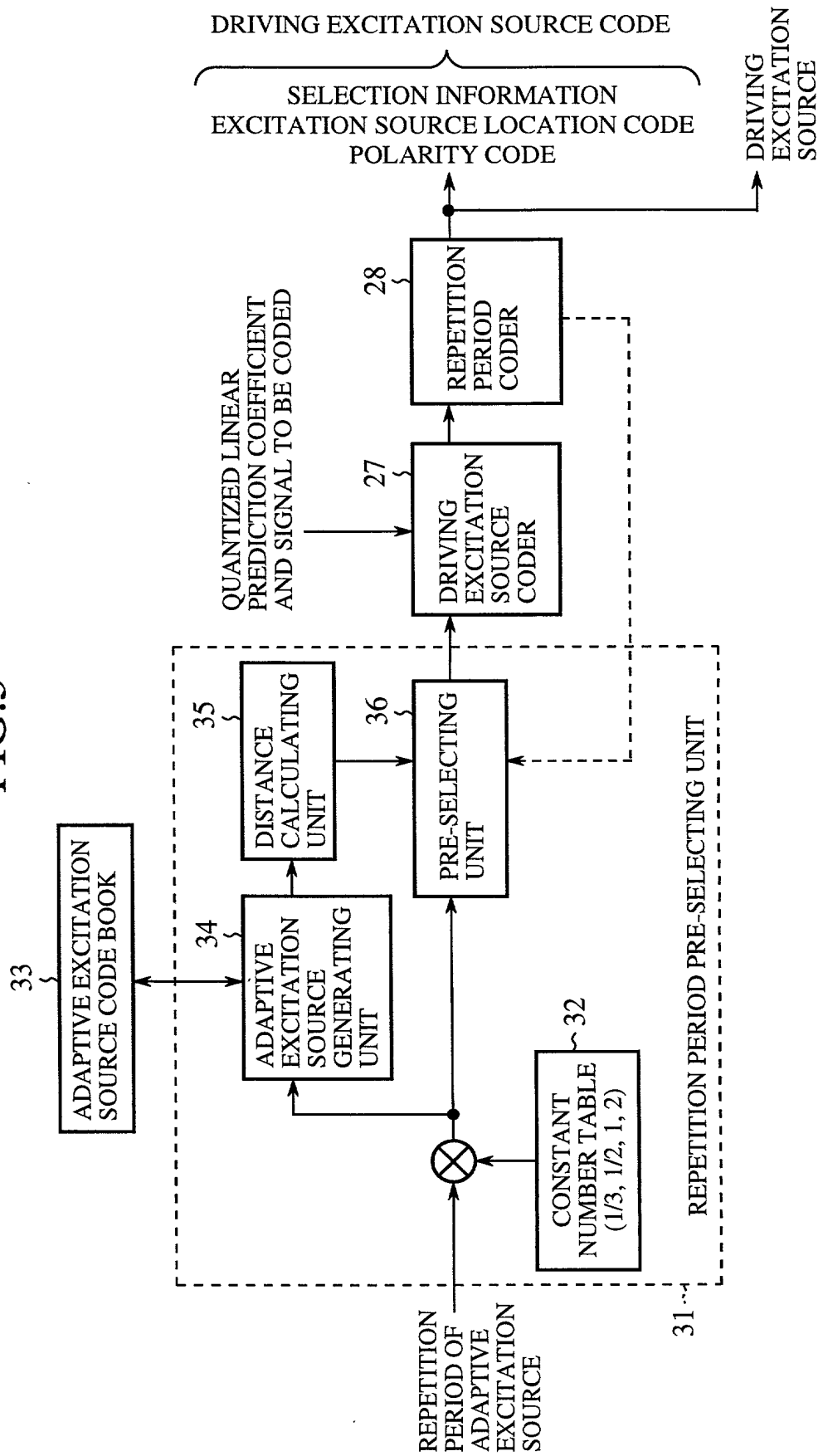


FIG. 6

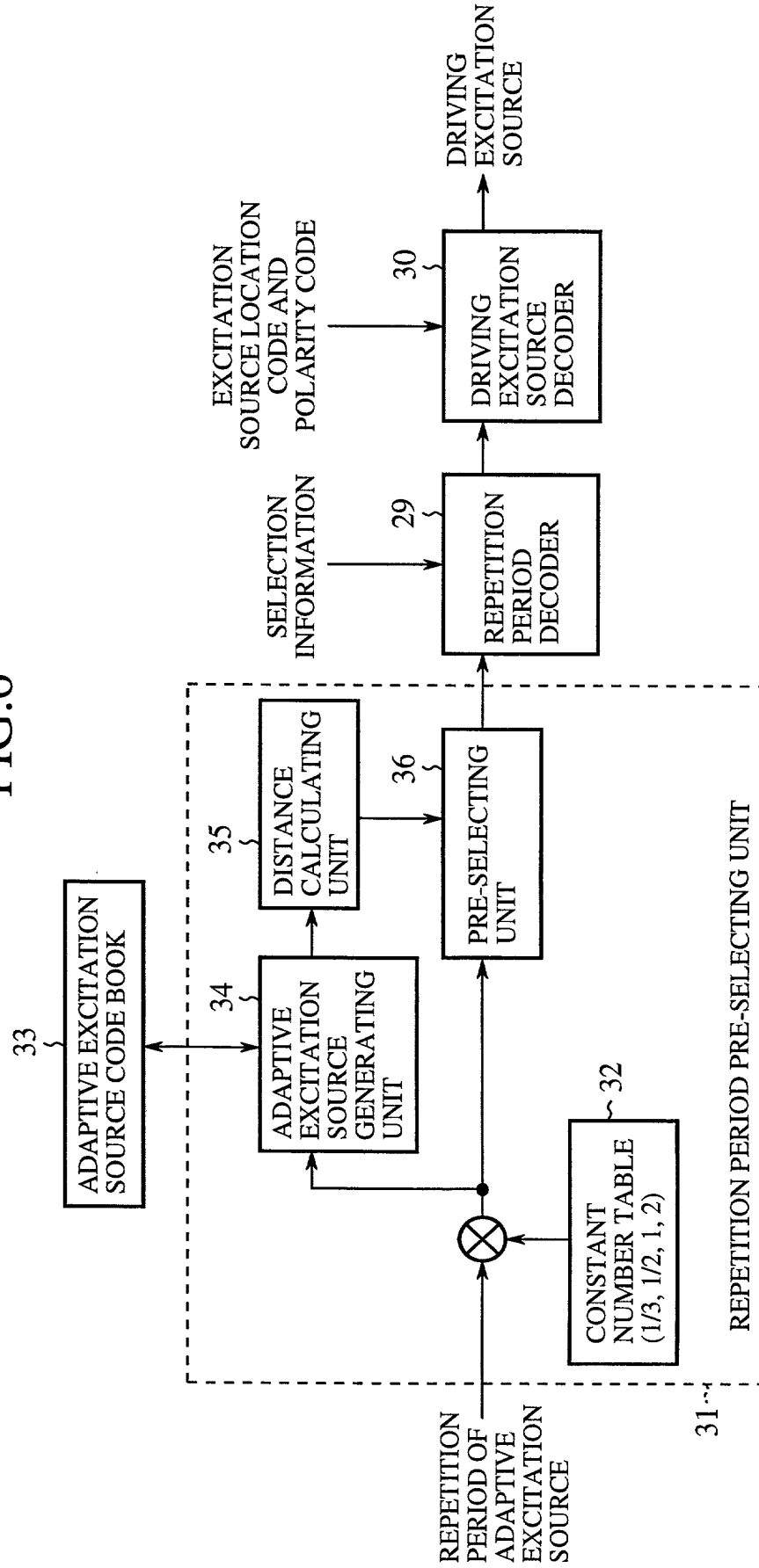
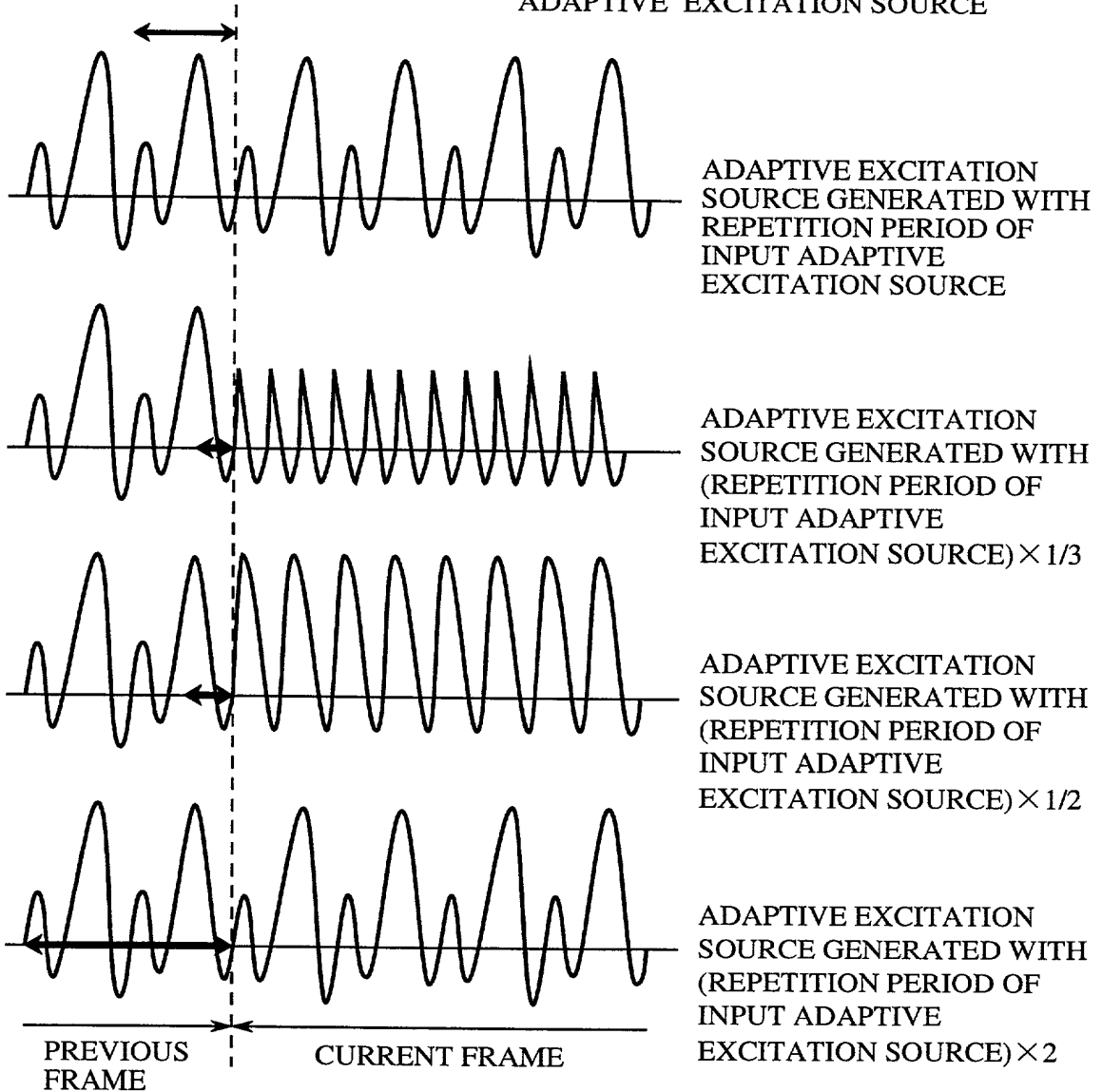
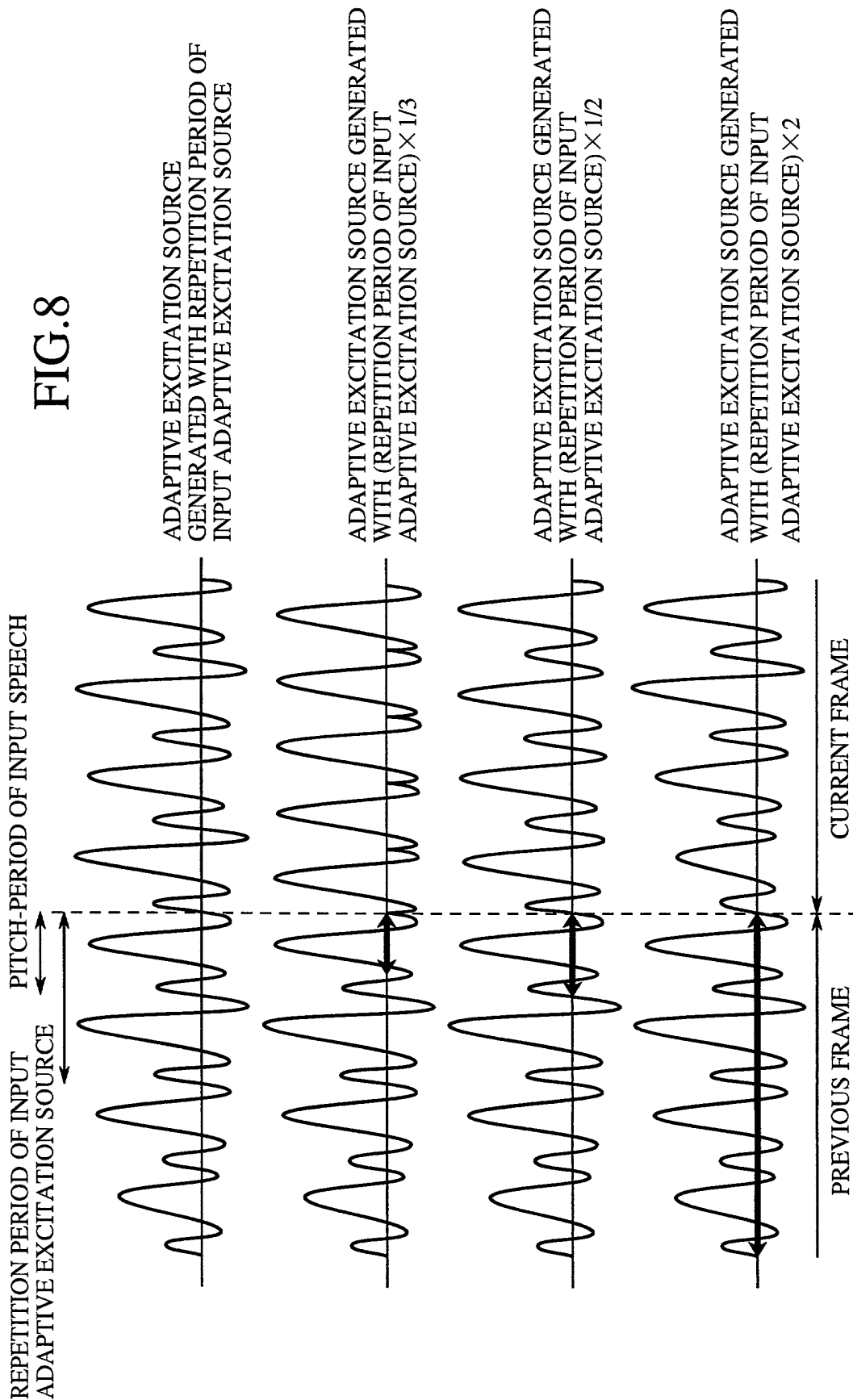


FIG.7

PITCH-PERIOD OF INPUT SPEECH = REPETITION PERIOD OF INPUT  
ADAPTIVE EXCITATION SOURCE







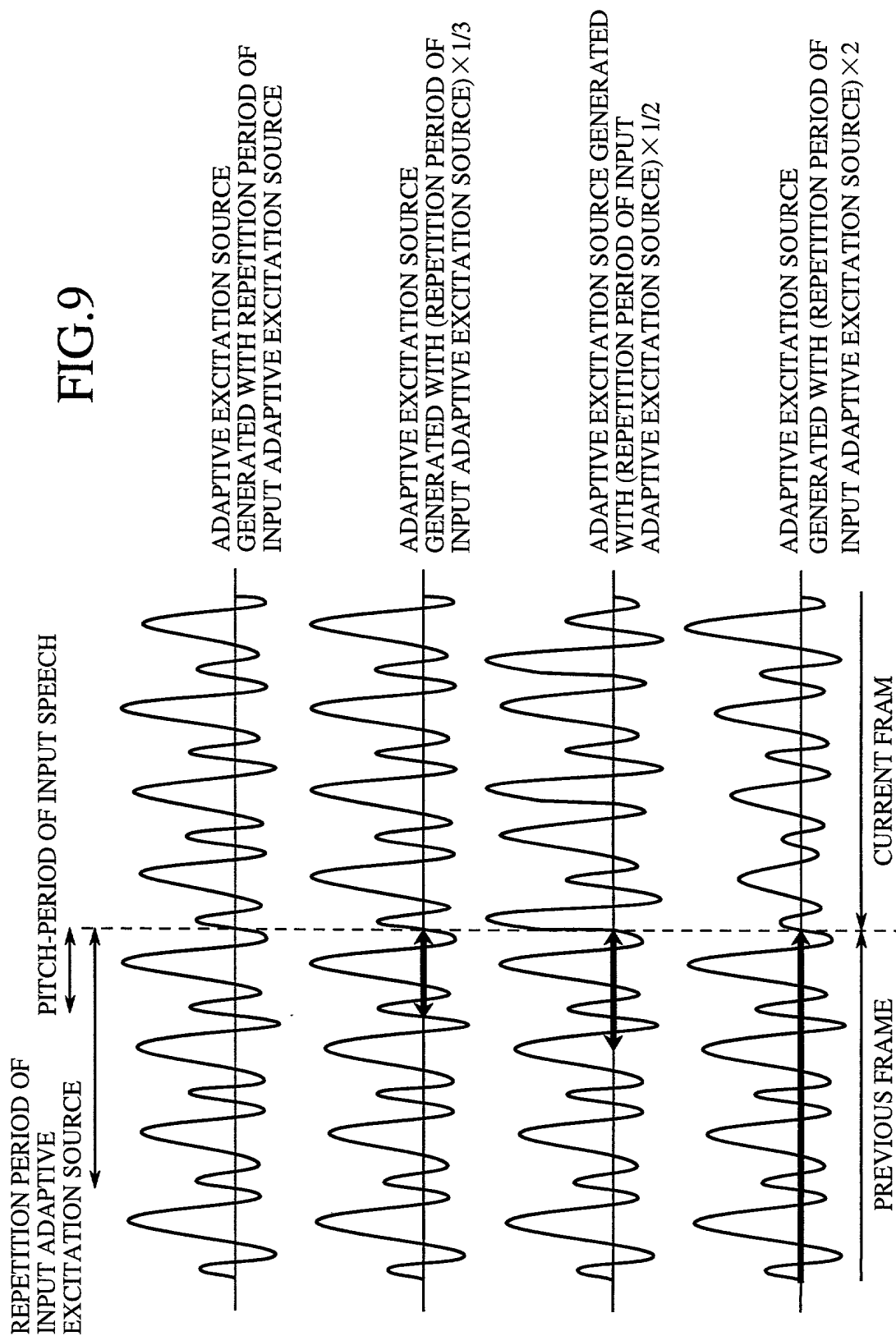


FIG.10

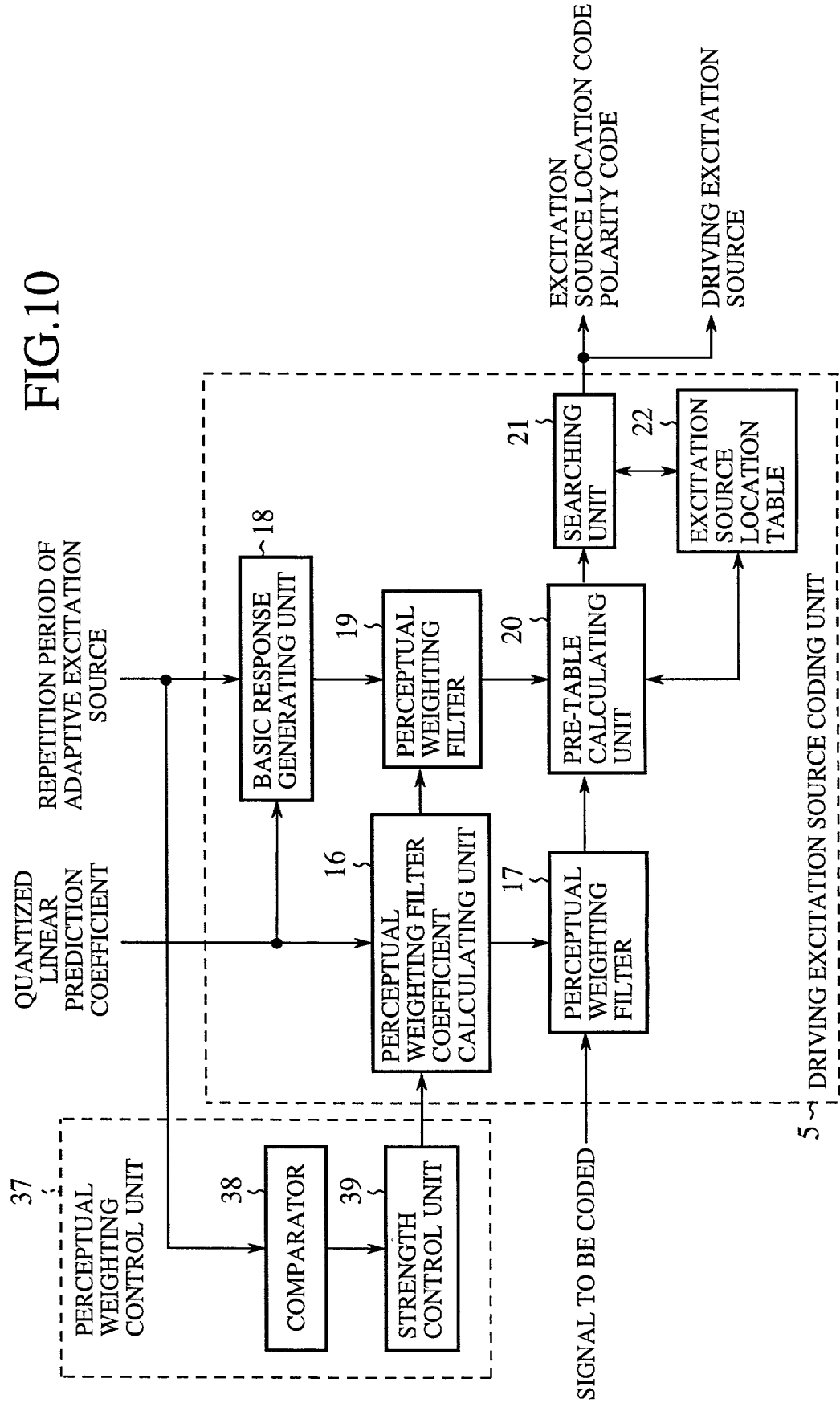




FIG.12

EXCITATION SOURCE LOCATION TABLE

EXCITATION SOURCE NUMBER	EXCITATION SOURCE LOCATION CANDIDATE	MAGNITUDE
1	0, 5, 10, 15, 20, 25, 30, 35	1.0
2	1, 6, 11, 16, 21, 26, 31, 36	1.0
3	2, 7, 12, 17, 22, 27, 32, 37	1.0
4	3, 8, 13, 18, 23, 28, 33, 38 4, 9, 14, 19, 24, 29, 34, 39	1.2

FIG.16 (PRIOR ART)

EXCITATION SOURCE LOCATION TABLE

EXCITATION SOURCE NUMBER	EXCITATION SOURCE LOCATION CANDIDATE
1	0, 5, 10, 15, 20, 25, 30, 35
2	1, 6, 11, 16, 21, 26, 31, 36
3	2, 7, 12, 17, 22, 27, 32, 37
4	3, 8, 13, 18, 23, 28, 33, 38 4, 9, 14, 19, 24, 29, 34, 39

FIG.13

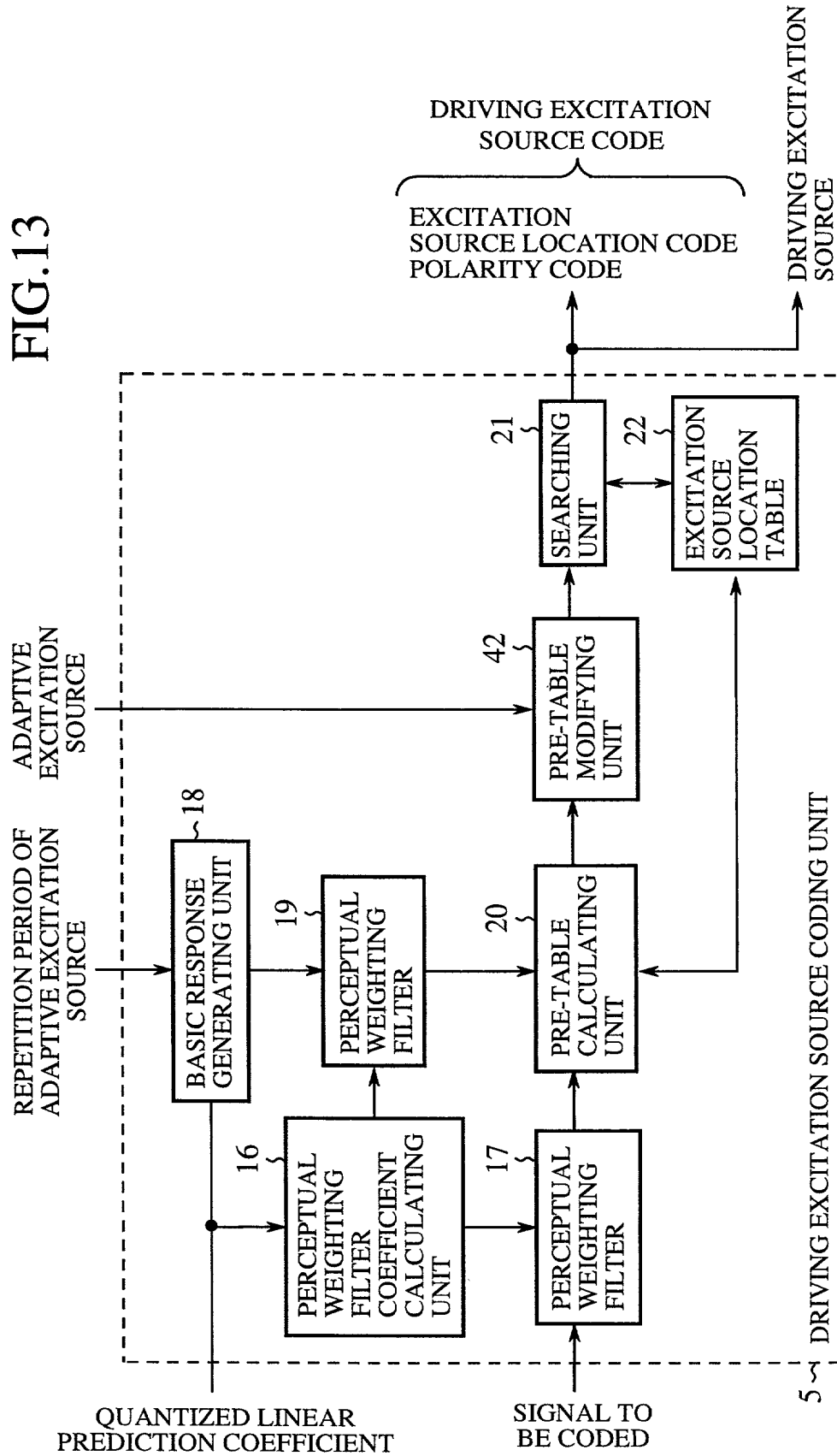


FIG.14

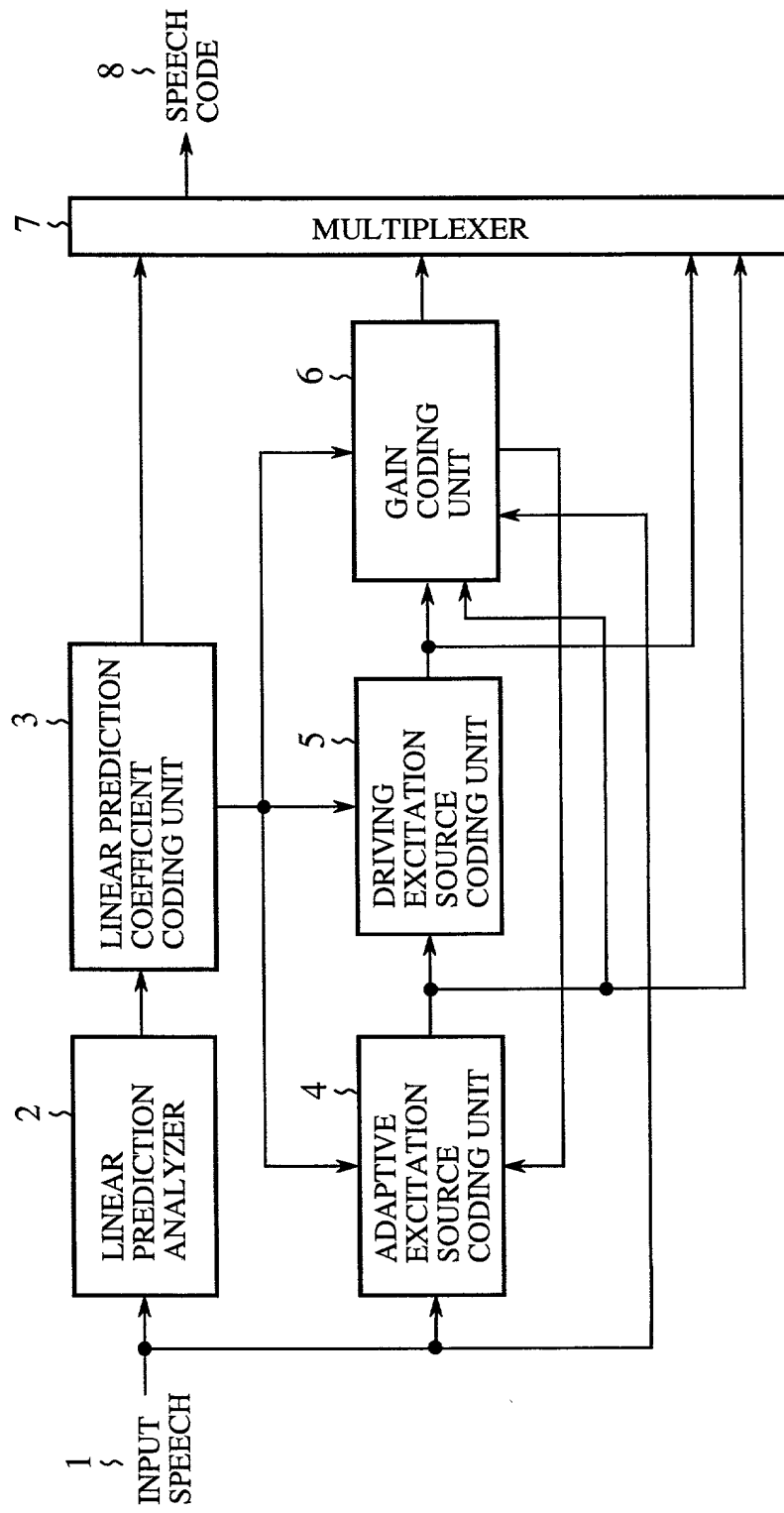


FIG.15

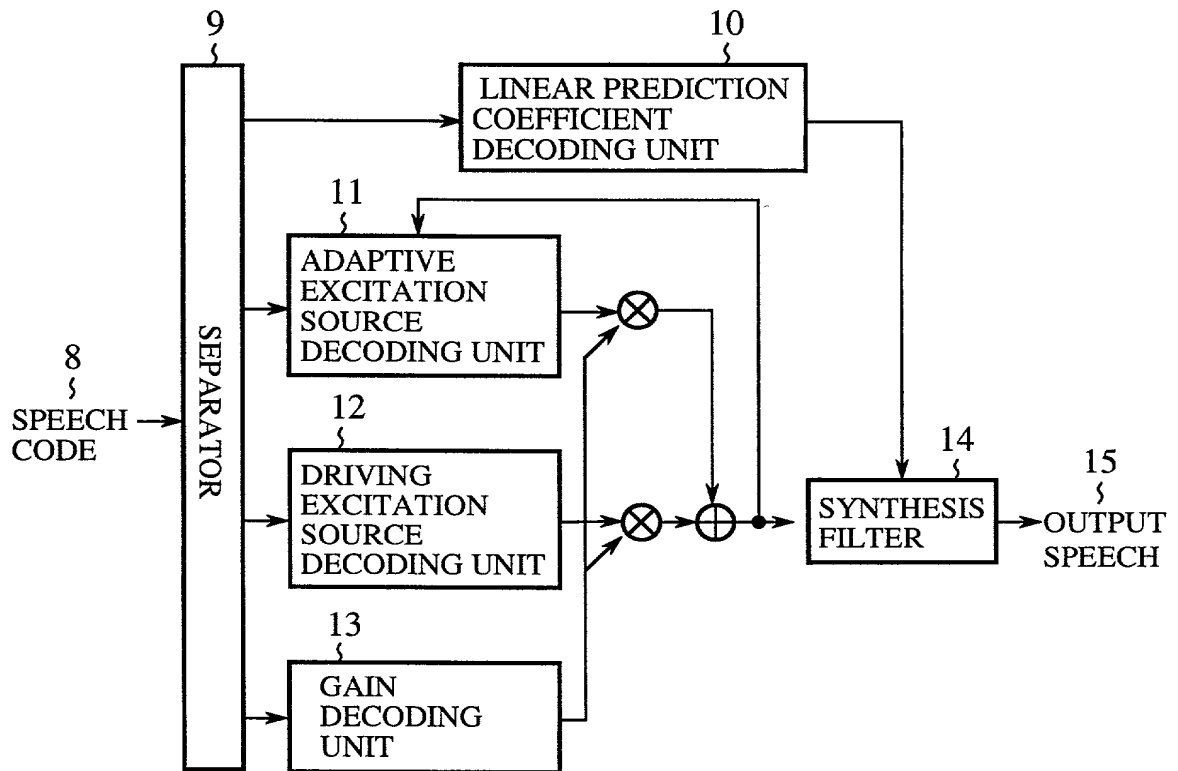




FIG.17  
(PRIOR ART)

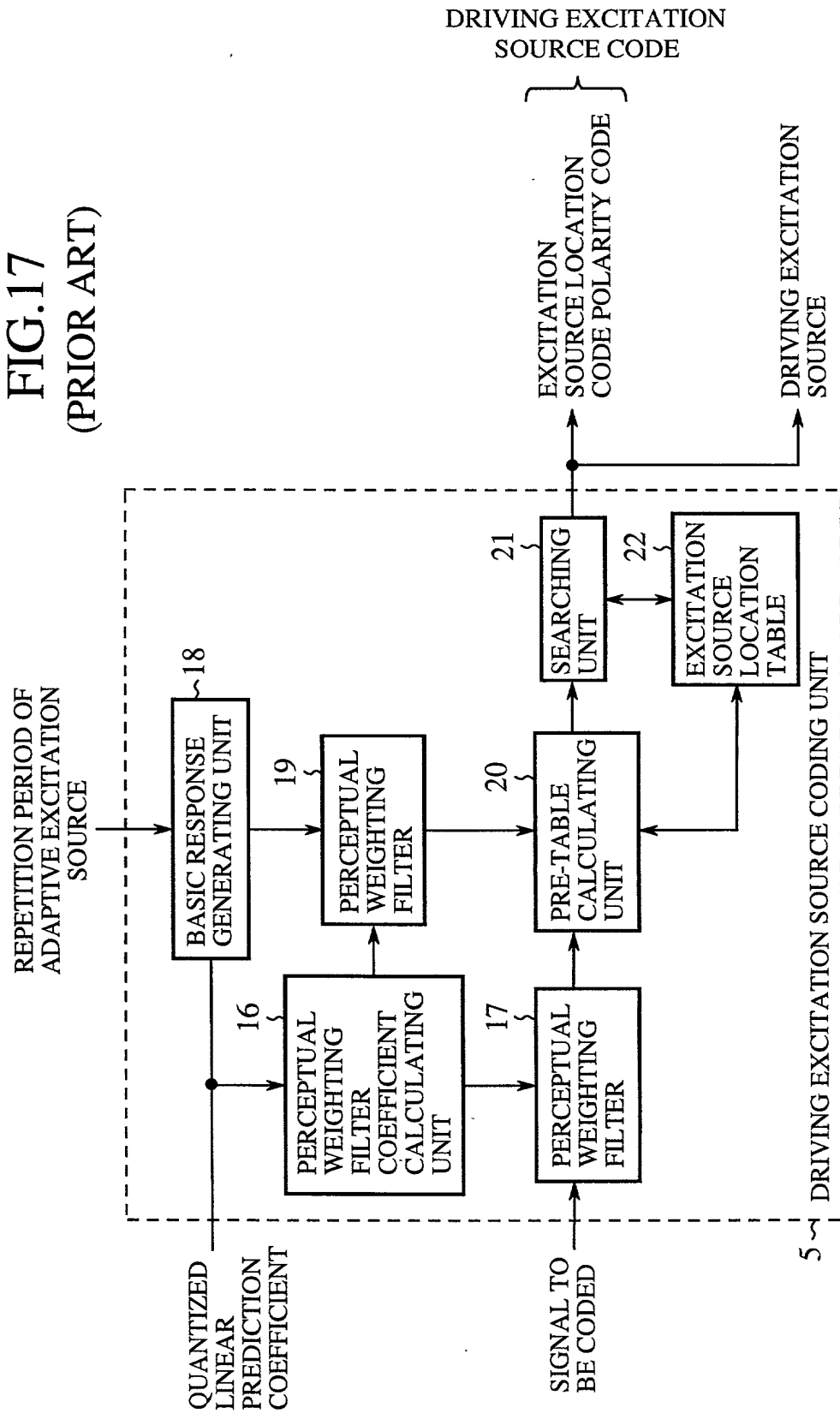


FIG.18 (PRIOR ART)

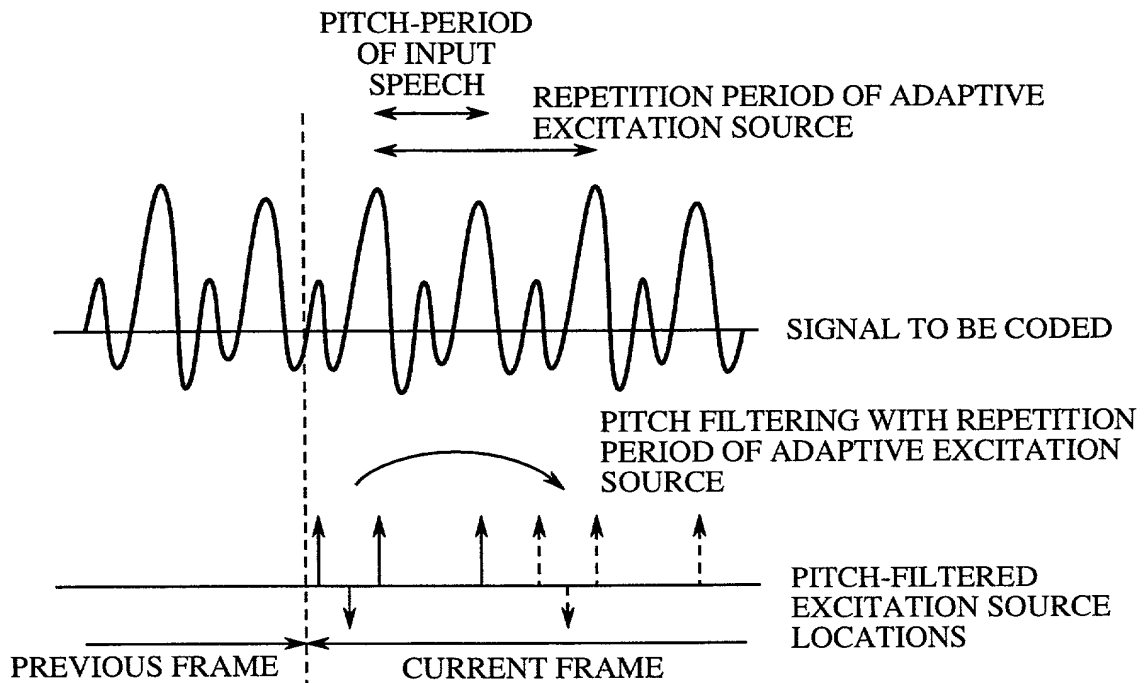
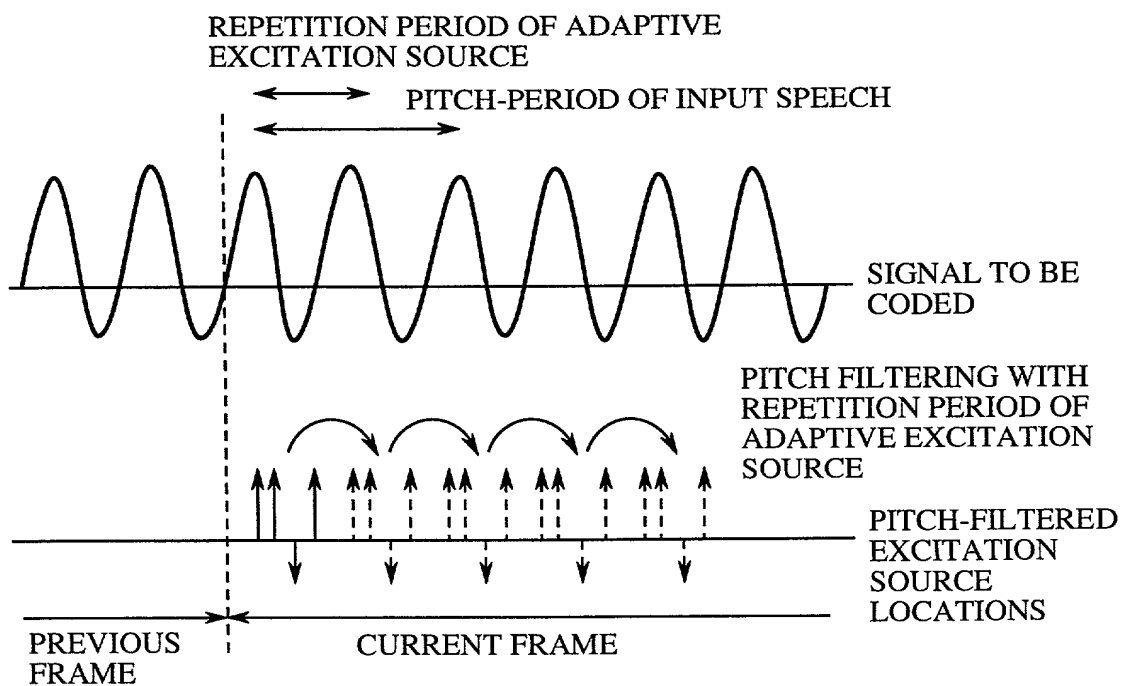


FIG.19 (PRIOR ART)



Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

## Declaration and Power of Attorney For Patent Application

1163-301P

### 特許出願宣言書及び委任状

### Japanese Language Declaration

### 日本語宣言書

下記の氏名の発明者として、私は以下の通り宣言します。

As a below named inventor, I hereby declare that:

私の住所、私書箱、国籍は下記の私の氏名の後に記載された通りです。

My residence, post office address and citizenship are as stated next to my name.

下記の名称の発明に関して請求範囲に記載され、特許出願している発明内容について、私が最初かつ唯一の発明者（下記の氏名が一つの場合）もしくは最初かつ共同発明者であると（下記の名称が複数の場合）信じています。

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled

"SPEECH CODING APPARATUS AND SPEECH

DECODING APPARATUS"

上記発明の明細書（下記の欄でx印がついていない場合は、本書に添付）は、

the specification of which is attached hereto unless the following box is checked:

☐ \_\_月\_\_日に提出され、米国出願番号または特許協定条約国際出願番号を\_\_\_\_とし、  
（該当する場合）\_\_\_\_に訂正されました。

☐ was filed on \_\_\_\_\_  
as United States Application Number or  
PCT International Application Number  
\_\_\_\_\_ and was amended on  
\_\_\_\_\_ (if applicable).

私は、特許請求範囲を含む上記訂正後の明細書を検討し、内容を理解していることをここに表明します。

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above.

私は、連邦規則法典第37編第1条56項に定義されるとおり、特許資格の有無について重要な情報を開示する義務があることを認めます。

I acknowledge the duty to disclose information which is material to patentability as defined in Title 37, Code of Federal Regulations, Section 1.56.

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

Japanese Language Declaration  
(日本語宣言書)

1163-301P

委任状: 私は下記の発明者として、本出願に関する一切の手続きを米特許商標局に対して遂行する弁理士または代理人として、下記の者を指名いたします。(弁理士、または代理人の氏名及び登録番号を明記のこと)

POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith (list name and registration number)

TERRELL C. BIRCH (Reg. No. 19,382)  
RAYMOND C. STEWART (Reg. No. 21,066)  
JOSEPH A. KOLASCH (Reg. No. 22,463)  
ANTHONY L. BIRCH (Reg. No. 26,122)

JAMES M. SLATTERY (Reg. No. 28,380)  
BERNARD L. SWEENEY (Reg. No. 24,448)  
MICHAEL K. MUTTER (Reg. No. 29,680)  
CHARLES GORENSTEIN (Reg. No. 29,271)

GERALD M. MURPHY (Reg. No. 28,977)  
LEONARD R. SVENSSON (Reg. No. 30,330)  
TERRY L. CLARK (Reg. No. 32,844)  
ANDREW D. MEIKLE (Reg. No. 32,868)

MARC S. WEINER (Reg. No. 32,181)  
ANDREW F. REISH (Reg. No. 33,443)  
JOE M. MUNCY (Reg. No. 32,334)  
C. JOSEPH FARACI (Reg. No. 32,350)

書類送付先

Send Correspondence to:

BIRCH, STEWART, KOLASCH & BIRCH, LLP  
P.O. BOX 747  
FALLS CHURCH, VA 22040-0747  
TEL: (703) 205-8000

直接電話連絡先: (名前及び電話番号)

Direct Telephone Calls to: (name and telephone number)

BIRCH, STEWART, KOLASCH & BIRCH, LLP  
TEL: (703) 205-8000

唯一または第一発明者名	Full name of sole or first inventor
	Hirohisa TASAKI
発明者の署名	Inventor's signature
日付	Date
	Hirohisa Tasaki October 20, 2000
住所	Residence
	Tokyo, Japan
国籍	Citizenship
	Japanese
私書箱	Post Office Address
	c/o MITSUBISHI DENKI KABUSHIKI KAISHA
	2-3, Marunouchi 2-chome, Chiyoda-ku, Tokyo 100-8310 Japan
第二共同発明者	Full name of second joint inventor, if any
	Tadashi YAMAURA
第二共同発明者	Second inventor's signature
日付	Date
	Tadashi Yamaura October 20, 2000
住所	Residence
	Tokyo, Japan
国籍	Citizenship
	Japanese
私書箱	Post Office Address
	c/o MITSUBISHI DENKI KABUSHIKI KAISHA
	2-3, Marunouchi 2-chome, Chiyoda-ku, Tokyo 100-8310 Japan

(第三以降の共同発明者についても同様に記載し、署名をすること)

(Supply similar information and signature for third and subsequent joint inventors.)

Japanese Language Declaration  
(日本語宣言書)

1163-301p

私は、米国法典第35編119条(a)-(d)項又は365条(b)項に基づき下記の、米国外の国の少なくとも一カ国を指定している特許協力条約365(a)項に基づく国際出願、又は外国での特許出願もしくは発明者証の出願についての外国優先権をここに主張するとともに、優先権を主張している、本出願の前に出願された特許または発明者証の外国出願を以下に、枠内をマークすることで、示しています。

Prior Foreign Application(s)

外国での先行出願

11-317205

(Number)  
(番号)

Japan

(Country)  
(国名)

(Number)  
(番号)

(Country)  
(国名)

私は、第35編米国法典119条(e)項に基づいて下記の米国外特許出願規定に記載された権利をここに主張いたします。

(Application No.)  
(出願番号)

(Filing Date)  
(出願日)

私は、下記の米国法典第35編120条に基づいて下記の米国外特許出願に記載された権利、又は米国外を指定している特許協力条約365条(c)に基づき権利をここに主張します。また、本出願の各請求範囲の内容が米国法典第35編112条第1項又は特許協力条約で規定された方法で先行する米国外特許出願に開示されていない限り、その先行米国外出願書提出日以降で本出願書の日本国内または特許協力条約国際提出日までの期間中に入手された、連邦規則法典第37編1条56項で定義された特許資格の有無に関する重要な情報について開示義務があることを認識しています。

(Application No.)  
(出願番号)

(Filing Date)  
(出願日)

(Application No.)  
(出願番号)

(Filing Date)  
(出願日)

私は、私自身の知識に基づいて本宣言書中で私が行なう表明が真実であり、かつ私の入手した情報と私の信じていることに基づき表明が全て真実であると信じていること、さらに故意になされた虚偽の表明及びそれと同等の行為は米国法典第18編第1001条に基づき、罰金または拘禁、もしくはその両方により処罰されること、そしてそのような故意による虚偽の声明を行えば、出願した、又は既に許可された特許の有効性が失われることを認識し、よってここに上記のごとく宣誓を致します。

I hereby claim foreign priority under Title 35, United States Code, Section 119 (a)-(d) or 365(b) of any foreign application(s) for patent or inventor's certificate, or 365(a) of any PCT International application which designated at least one country other than the United States, listed below and have also identified below, by checking the box, any foreign application for patent or inventor's certificate, or PCT International application having a filing date before that of the application on which priority is claimed.

Priority Not Claimed  
優先権主張なし

8/November/1999

(Day/Month/Year Filed)  
(出願年月日)

(Day/Month/Year Filed)  
(出願年月日)

I hereby claim the benefit under Title 35, United States Code, Section 119(e) of any United States provisional application(s) listed below.

(Application No.)  
(出願番号)

(Filing Date)  
(出願日)

I hereby claim the benefit under Title 35, United States Code, Section 120 of any United States application(s), or 365(c) of any PCT International application designating the United States, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT International application in the manner provided by the first paragraph of Title 35, United States Code Section 112, I acknowledge the duty to disclose information which is material to patentability as defined in Title 37, Code of Federal Regulations, Section 1.56 which became available between the filing date of the prior application and the national or PCT International filing date of application.

(Status: Patented, Pending, Abandoned)  
(現況: 特許許可済、係属中、放棄済)

(Status: Patented, Pending, Abandoned)  
(現況: 特許許可済、係属中、放棄済)

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

PRIOR ART DOCUMENTS

1. JP-A 10-232696  
Publication Date: September 2, 1998
2. JP-A 10-312198  
Publication Date: November 24, 1998
3. "Basic Algorithm of Conjugate-Structure Algebraic CELP (CS-ACELP) Speech Coder" by Akitoshi Kataoka, Shinji Hayashi, Takehiro Moriya, Sachiko Kurihara and Kazunori Mano, NTT R&D Vol. 45 No. 4, 1996
4. "Improved CELP speech coding using adaptive pulse position algebraic codebook" by Katsumi Tsuchiya, Tadashi Amada and Kimio Miseki, Kansai Research Lab. Toshiba, March 1999

TRANSLATION OF THE RELEVANT PART

3. "Basic Algorithm of Conjugate-Structure Algebraic CELP (CS-ACELP) Speech Coder"

ABSTRACT

After lengthy discussion of standardization of an 8Kbit/s speech coding algorithm that will be used for the Future Public Land Mobile Telecommunication System (FPLMTS), the ITU-T approved the speech coder CS-ACELP (Conjugate-Structure Algebraic CELP) in November 1995 meeting. This paper overviews of CS-ACELP and discusses the algebraic codebook, perceptual weighting, and quantization methods for LSP parameters and gain. It also presents the results of subjective quality tests.

4. "Improved CELP speech coding using adaptive pulse position algebraic codebook"

3.1 Consideration of Pitch filter

in a section where a pitch filter process is performed with